

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 6 :</b> C07H 21/00, C12N 1/15, 1/21, 5/10, 15/11, 15/63		<b>A1</b>	<b>(11) International Publication Number:</b> WO 99/31117 <b>(43) International Publication Date:</b> 24 June 1999 (24.06.99)																																						
<b>(21) International Application Number:</b> PCT/US98/27059 <b>(22) International Filing Date:</b> 17 December 1998 (17.12.98)			<b>20874 (US).</b> RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). SHI, Yang-gu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). KYAW, Hla [MM/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). WEI, Ying-Fei [CN/US]; 1714-C Marina Court, San Mateo, CA 94403 (US). FLORENCE, Kimberly [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). DUAN, Roxanne, D. [US/US]; 5515 Northfield Road, Bethesda, MD 20817 (US). FLORENCE, Charles [US/US]; 12805 Altantic Avenue, Rockville, MD 20851 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 120 Fox Run Drive, Tewksbury, MA 01876 (US). YU, Guo-Liang [CN/US]; 1714-C Marina Court, San Mateo, CA 94403 (US). JANAT, Fouad [SY/US]; 140 High Street, No. 202, Westerly, RI 02891 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US).  <b>(74) Agents:</b> BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).  <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> With international search report.																																						
<b>(30) Priority Data:</b> <table><tr><td>60/070,923</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,007</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,057</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,006</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,008</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,054</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,064</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,053</td><td>18 December 1997 (18.12.97)</td><td>US</td></tr><tr><td>60/068,169</td><td>19 December 1997 (19.12.97)</td><td>US</td></tr><tr><td>60/068,368</td><td>19 December 1997 (19.12.97)</td><td>US</td></tr><tr><td>60/068,367</td><td>19 December 1997 (19.12.97)</td><td>US</td></tr><tr><td>60/068,369</td><td>19 December 1997 (19.12.97)</td><td>US</td></tr><tr><td>60/068,365</td><td>19 December 1997 (19.12.97)</td><td>US</td></tr></table> <b>(71) Applicant (for all designated States except US):</b> HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> MOORE, Paul, A. [US/US]; 19005 Leatherbark Drive, Germantown, MD				60/070,923	18 December 1997 (18.12.97)	US	60/068,007	18 December 1997 (18.12.97)	US	60/068,057	18 December 1997 (18.12.97)	US	60/068,006	18 December 1997 (18.12.97)	US	60/068,008	18 December 1997 (18.12.97)	US	60/068,054	18 December 1997 (18.12.97)	US	60/068,064	18 December 1997 (18.12.97)	US	60/068,053	18 December 1997 (18.12.97)	US	60/068,169	19 December 1997 (19.12.97)	US	60/068,368	19 December 1997 (19.12.97)	US	60/068,367	19 December 1997 (19.12.97)	US	60/068,369	19 December 1997 (19.12.97)	US	60/068,365	19 December 1997 (19.12.97)
60/070,923	18 December 1997 (18.12.97)	US																																							
60/068,007	18 December 1997 (18.12.97)	US																																							
60/068,057	18 December 1997 (18.12.97)	US																																							
60/068,006	18 December 1997 (18.12.97)	US																																							
60/068,008	18 December 1997 (18.12.97)	US																																							
60/068,054	18 December 1997 (18.12.97)	US																																							
60/068,064	18 December 1997 (18.12.97)	US																																							
60/068,053	18 December 1997 (18.12.97)	US																																							
60/068,169	19 December 1997 (19.12.97)	US																																							
60/068,368	19 December 1997 (19.12.97)	US																																							
60/068,367	19 December 1997 (19.12.97)	US																																							
60/068,369	19 December 1997 (19.12.97)	US																																							
60/068,365	19 December 1997 (19.12.97)	US																																							
<b>(54) Title:</b> 110 HUMAN SECRETED PROTEINS																																									
<b>(57) Abstract</b>  The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.																																									

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Licchtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## 110 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and  
5 their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or  
10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum  
15 (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the  
20 extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or  
25 secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include  
30 the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using  
35 secreted proteins or the genes that encode them.

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

### *Detailed Description*

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig



analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard,  
5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained  
10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the  
15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages  
20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even  
25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include  
30 Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such  
35 as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5       The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and  
10   double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability  
15   or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

      The polypeptide of the present invention can be composed of amino acids joined  
20   to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs,  
25   as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be  
30   branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a  
35   nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## **Polynucleotides and Polypeptides of the Invention**

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 1**

The translation product of this gene shares sequence homology with a neurogenic secreted signaling protein, in addition to the human UDP-galactose:2-acetamido-2-deoxy-D-glucose3beta-galactosyltransferase (See Genbank Accession No. gnllPIDle1237254 ) which is thought to be vital in glycoprotein biosynthesis. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GLGPAQVALSLQGPA (SEQ ID NO:239), SSWMAGTQPRTSWWEMSS AKPCPTGTLRSNTSSHPQCTGPPTTHPMLVGEDMSCPEPQCGASRLSWKMNS

SPLMMSLWVCA (SEQ ID NO:240), QPRTSWWEMSSAKPCPTGTLRSN (SEQ ID NO:241), MSCPEPQCGASRLSWKMLNSSPL (SEQ ID NO:242), WVALYIEG GMKYLTLVFLLGRAWRMTPTRRSWAGSQPSRNSNTLGTWTKTSSSPFSMK WAWGQAATTQRCRCSSLSVRLKKSSVKSHWRMSSNSLLS (SEQ ID NO:243),  
 5 GGMKYLTLVFLLGRAWRMSTS (SEQ ID NO:244), SQPSRNSNTLGTWTKTS SSPFSMKW (SEQ ID NO:245), and/or TTQRCRCSSLSVRLKKSSVKSHWRMS (SEQ ID NO:246). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on  
 10 chromosome 12. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 12.

This gene is expressed primarily in human fetal brain, epileptic frontal cortex and 12 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
 15 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural or immune disorders, particularly neurodegenerative conditions such as epilepsy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the  
 20 tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell  
 25 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:125 as residues: Ala-27 to Ser-38, Pro-43 to Asn-54, Thr-115 to Asp-121, Leu-225 to Val-232, Pro-247 to Gly-252, Arg-306 to Leu-311.

30 The tissue distribution in fetal brain tissue, combined with the homology to a neurogenic secreted signaling protein, in addition to the conserved UDP-galactose:2-acetamido-2-deoxy-D-glucose3beta-galactosyltransferase protein indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, detection, and/or prevention of a variety of neural disorders, particularly  
 35 epilepsy and other disorders of the CNS. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are

not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. In addition, the protein may show utility in the creation of novel therapeutics which depend upon the localizing benefits (cell and tissue specificity) of glycoproteins. This protein may also be used to produce physiologically active saccharide chains and variants, and for improvement of saccharide chains bound to physiologically active proteins. Expression within fetal tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:11 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1257 of SEQ ID NO:11, b is an integer of 15 to 1271, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:11, and where b is greater than or equal to a + 14.

## 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 2

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: ASTLAQ

TTGTCKXXXSSRRARSRTQRXFQLRPDKRSAPSLQFIQAQEELSKENTGRQLA  
 AREAVLALEGSTQLTGPVTQVAASKTHCSGMALTASVPVVLGAAPAKXPTQ  
 5 NXPGQXGRAXXKVXTSWXXVATKVLHGLEVSTHLGKRKLSGRSWLPGP  
 ALHATPSQSHTQTGSQIVHPPQGEVREVGRGRGQPPAQPVHAHPSQQHPSPAH  
 LAGLSLWTGTA (SEQ ID NO:247), AMLETWRPGPSXGELATNSGQRASQDSQ  
 HSPPHVRAHLLISPLPAFPSMGGPAGRSAPXXLTETKSELQRLRRRQARASXS  
 XPAGEPGAGHSDSFNCVPTNGQPLRSCSLSKLRRSFLKRTQGDSWLPEKQSW  
 10 LWKAPPS (SEQ ID NO:248), SHQSHLINPASSAKGSAQLKAQPPAHVLGGT  
 GQEGPPPTADQPESPGWDPSSFTNGSSGPRALPTS VHPTLQQGAPCRRNWA  
 PCRGLVETRMLRRQLPHGTSKRDLGWASLQRGSPQETPQ (SEQ ID NO:249),  
 RPDKRSAPSLQFIQAQEELSKENTGRQLAAREAV (SEQ ID NO:250), ATPSQ  
 SHTQTGSQIVHPPQGEVREVGRGRGQPP (SEQ ID NO:251), QDSQHSPPHVR  
 15 AHLLISPLPAFPSMGGPA (SEQ ID NO:252), DSFNCVPTNGQPLRSCSL  
 KLRRSFLKR (SEQ ID NO:253), KGSWAQLKAQPPAHVLGGTGQEGPP (SEQ ID  
 NO 254:), KPSHQ (SEQ ID NO:256), and/or APSLLQFIQAQEELSKENTGRQLA  
 AR (SEQ ID NO:255). Polynucleotides encoding these polypeptides are also  
 encompassed by the invention.

20 This gene is expressed exclusively in adult human testis.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, reproductive disorders, particularly abnormalities of the testis, in addition  
 25 to impotence and infertility. Similarly, polypeptides and antibodies directed to these  
 polypeptides are useful in providing immunological probes for differential identification  
 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
 particularly of the male reproductive system, expression of this gene at significantly  
 higher or lower levels may be routinely detected in certain tissues or cell types (e.g.,  
 30 reproductive, testicular, adrogen regulated, and cancerous and wounded tissues) or  
 bodily fluids (e.g., lymph, serum, plasma, seminal fluid, urine, synovial fluid and  
 spinal fluid) or another tissue or cell sample taken from an individual having such a  
 disorder, relative to the standard gene expression level, i.e., the expression level in  
 healthy tissue or bodily fluid from an individual not having the disorder.

35 Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO:126 as residues: His-45 to Gly-56, Trp-62 to Tyr-68, His-94 to Trp-100.

The tissue distribution in testis indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention for abnormalities of the reproductive system. In addition, expression of this gene product in the testis may implicate this gene product in normal testicular function. This gene product may be useful in the treatment of male infertility, and/or could be used as a male contraceptive. Moreover, the protein product of this gene may be useful in the treatment, detection, and/or prevention of a variety of disorders related to androgen-regulated tissues, particularly the prostate gland. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:12 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1437 of SEQ ID NO:12, b is an integer of 15 to 1451, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:12, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 3

The translation product of this gene shares sequence homology with the human VAKTI precursor (See Genbank Accession No. gnllPIDle1311078 (AJ228139)), in addition to the ovoinhibitor and thrombin inhibitors, which are thought to be important in inhibition of protease activities. Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of monocytes to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both immune cells, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocytes. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to

receptors of a particular cell. Moreover, when tested against NIH3T3 and U937 cell lines, supernatants removed from cells containing this gene activated the EGR1 (early growth response) and GAS (gamma activating sequence) promoter elements. Thus, it is likely that this gene activates fibroblasts or hematopoietic cells through the EGR1 and/or

5 JAK-STAT signal transduction pathway. EGR1 is a separate signal transduction pathway from Jak-STAT, genes containing the EGR1 promoter are induced in various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation. GAS is also a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal

10 transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: CSYRPQFPVDPRVRATCIVFN (SEQ ID NO:257), GTENLLA

15 PERTILSRAQMKGCMATPAPCVRSSSKQKKKKRKRKRVXQETKDNLRVQLPL XSCVVNXANPGKTDGFFAPERMTSPRAQMEKCMATPAPCVRPSFNKKKEQE QRLKEKLQRKSAVNFGTK (SEQ ID NO:258), LLAPERTILSRAQMKGCMAT PAPCVR (SEQ ID NO:259), PGKTDGFFAPERMTSPRAQMEKCM (SEQ ID NO:260), EQRLKEKLQRKSAVNFG (SEQ ID NO:261), KTLLENFSTQGTFFVAMH

20 PAVRATDWITLPCTKKPSISHLFFXFLAKILFSISSNSSFTLSLGIFSFFXXQLST HCTLIAMRLPIRTKNRIIFPCASKSSISNKGPKSTAYILLWITALTFPFTFYTNL GPGFRILSTQCTSVVICFPICATNSFIIRTDKIPISFSFFKIITIQLC WGSSLGSSC (SEQ ID NO:262), MHPAVRATDWITLPCTKKPSIS (SEQ ID NO:263), LIAMRLP IRTKNRIIFP (SEQ ID NO:264), SSISNKGPKSTAYILL WITALTFPFT (SEQ ID NO:265), IIIRTDKIPISFSFFKIITIQLC (SEQ ID NO:266), NDGQCLAYNTTHY

25 RERAMTSHARVSLGPSRDPLERPPFFFFFFFFFFFFKFEHTGTHGTLVSMHFAI WATDRIMLPGAYKCSIPHLVPKFTADFLCSFSFSLCSCSFLLKEGLTHGAGVA MHFSIWALDGVILSGAKKPSVFPGFAXFTTQLXKGSCITL RLSFVS (SEQ ID NO:267), CLAYNTTHYRERAMTSHARVSL (SEQ ID NO:268), GTLVSMHFAI

30 WATDRIMLPGAYKCSIPHLVP (SEQ ID NO:269), GVILSGAKK PSVFPGFAX FTTQLX (SEQ ID NO:270), KKASHMEQVLPCIFPSGPWMGSFSLXQKSRPF FLDLRXSLHNSXKEAVLLDCLLFLXXPSFFFFSSSSAWKKTSHMEQVLPCIT FPSGPWIGLFSLVQASFPFLTSFRYSLQSSAYEVAFPDSSLFLARASAFFSSSFA WK (SEQ ID NO:271), CIFPSGPWMGSFSLXQKSRPFLLDLRXS (SEQ ID

35 NO:272), WIGLFSLVQASFPFLTSFRYSLQSSAYE (SEQ ID NO:273), NSAVN IKIRQRMEYFSVPEKMTLFFVQMKGCMATCVPCVKPTSKQKMKKRKRLKHE LETKENLEKQPHMQSFAVNIESL (SEQ ID NO:274), IKIRQRMEYFSVPEKMTL



FVVQM (SEQ ID NO:275), and/or VKPTSKQKMKRRLKHELETKENL (SEQ ID NO:276). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5.

This gene is expressed primarily in heart, tonsils, Hodgkin's lymphoma, neuroblastoma, leukocyte and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cardiovascular, immune, or hemodynamic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., cardiovascular, muscle, immune, hematopoietic, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, pulmonary surfactant or sputum, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:127 as residues: Ala-20 to Gln-27.

The tissue distribution in heart and immune cells and tissues, the homology to protease inhibitors, in addition to the detected calcium flux, EGR1, and GAS biological activities indicates polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hemodynamic or vascular disorders, including hemorrhage, heart failure, and embolism, because proteases and their inhibitors are often involved in the cascades controlling hemodynamic controls. Protein may also show utility in the treatment, detection, and/or prevention of a variety of metabolic (i.e. cellular or physiological) and/or proliferative disorders in which aberrant regulation of a protease is thought to be involved, particularly in the premature activation of zymogens, for example. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines;

- immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.
- Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:13 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.
- Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2303 of SEQ ID NO:13, b is an integer of 15 to 2317, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:13, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 4

- The translation product of this gene shares sequence homology with the ecotropic retrovirus receptor and the human cationic amino acid transporter-3 (See Genbank Accession No. gnlIPID1198517) which are thought to be important in viral infections and amino acid and polyamine transport. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:
- PRVRGTVVRLRQHRPSAYILVSTVLTLMPVWHSLDPDSALADAFYQRGYRWAG  
FIVAAGSICA (SEQ ID NO:277), TVVRLRQHRPSAYILVSTVLTLMPV (SEQ ID  
NO:278), WHSLDPDSALADAFYQRGYRWAGFIV (SEQ ID NO:279), TPSCSASS  
SPCHALSMPWPPMGSSSRCLPMCTPGHRCLWRAPWRSGSSRPSWHCCWTWS

RWFSSCPLAHSWPTHSWPPVSLCCASRSLPRPAPQAQPALAP (SEQ ID NO:280), LSMPWPPMGSSSRCLPMCTPGHRC (SEQ ID NO:281), APWRSGSS RPSWHCCWTWSRWFSSCPL (SEQ ID NO:282), THSWPPVSLCCASRSL PRPAPQ (SEQ ID NO:283), AYILVSTVLTLMPVWHSLDPDSALADAFYQRGYRW  
 5 AGFIVAAGSICAMNTVLLSLLFSLP (SEQ ID NO:284), PWHSLDPDSALADAF YQRGYRWAGFIVAAGS (SEQ ID NO:285), RIVYAMAADGLFFQVFAHVHPRTQ VPV (SEQ ID NO:286), DLESLVQFLSLGTLLA (SEQ ID NO:287), YTFVATSII VLRQK (SEQ ID NO:288), LTKQQSSFSDHLQLVGTVHASVPEPGELKPA (SEQ ID NO:289), LRPYLGFLDGYSPGAVVTWALGVMLASAITIGCVLVFGNSTL  
 10 HLPHWGYI (SEQ ID NO:290), PGAVVTWALGVMLASAITIGCVLVFGN (SEQ ID NO:291), GAHQQQYREDLFQIPMVPLIPALSIVLNICLMLKLSYLTWVRFSTW LLMGLAV (SEQ ID NO:292), MVPLIPALSIVLNICLMLKLSYLTWV (SEQ ID NO:293), and/or YFGYGIRHSKENQRELPLNSTHYVVFP (SEQ ID NO:294).  
 Polynucleotides encoding these polypeptides are also encompassed by the invention.

15 This gene is expressed primarily in placenta and brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, neural, or metabolic disorders, in addition to viral  
 20 infections. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system and placenta, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g.,  
 25 reproductive, neural, hepatic, metabolic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, bile, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:128 as residues: Gln-87 to Ser-99, Pro-102 to Phe-110, Gln-204 to Leu-211, Ser-262 to Glu-268, Pro-294 to His-305.

The tissue distribution in placenta, combined with the homology to a retroviral receptor and cationic amino acid transporters, indicates polynucleotides and  
 35 polypeptides corresponding to this gene are useful for the diagnosis and intervention of viral infections, or diseases and malfunctions related to amino acid transport. Specifically, soluble forms of this protein or, polynucleotides of the present invention,

can be used to bind to retroviruses so as to prevent their entry and infection of susceptible cells. They can be used for therapy/prevention of HIV infection and certain forms of leukaemia. Polynucleotide or polypeptides of the present invention can be used to identify susceptibility to retroviral infection. Based upon the tissue distribution in the brain, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:14 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1458 of SEQ ID NO:14, b is an integer of 15 to 1472, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:14, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 5

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: FPPSPAPPHSLPLRSWLWSRQMG (SEQ ID NO:295), GTSF RGMISTQPGSTPLASFKILALESADGHGGCSAGNDIGPYGERDDQQVFIQKVVP SASQLFVRLSSTGQRVCSVRSVDGSPTTAFTVLECEGSPAARLSAPALPAHWP

GQRQLGHVGPNNHRHGRPRPGPCRWPDGAR ADGTAGTL (SEQ ID NO:296),  
 PGSTPLASFKILALESADGHGGCSAGNDI (SEQ ID NO:297), GERDDQQVF  
 IQKVVPSASQLFVRL (SEQ ID NO:298), RSVDGSPTTAFTVLECEGSPAARLS  
 (SEQ ID NO:299), PALPAHWPGQRQLGHVGPNNHRHGRPR (SEQ ID NO:300),  
 5 PFIPRRPWPEPGVPTGIREAPESPRTRASQGIMAAALFKKEVSLSFILGGVVRG  
 VPRPLEGHGAGVGGRRRSGLRTSSWQRSTKLPPRRRRASACGGLGLPRWP  
 DKEVLLEAEWRLVREMRGEGGLGRQPHEGAERSRRGQLTVFQLFHQLLLRQATC  
 RGLA EAVHGGG (SEQ ID NO:301), PGVPTGIREAPESPRTRASQGIMAAALF  
 KKEV (SEQ ID NO:302), FILGGVVRGVPRLPLEGHGAGVGGRRRSGLP (SEQ ID  
 10 NO:303), GLPRWPDKEVLLEAEWRLVREMRG (SEQ ID NO:304), and/or GAER  
 SRRGQLTVFQLFHQLLLRQ (SEQ ID NO:305). Polynucleotides encoding these  
 polypeptides are also encompassed by the invention.

This gene is expressed primarily in human fetal kidney, and to a lesser extent, in  
 thymus and bone marrow cell line (RS4;11).

15 Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, developmental, metabolic, immune or hematopoietic disorders.  
 Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
 20 providing immunological probes for differential identification of the tissue(s) or cell  
 type(s). For a number of disorders of the above tissues or cells, particularly of the  
 immune system, expression of this gene at significantly higher or lower levels may be  
 routinely detected in certain tissues or cell types (e.g., developmental, metabolic,  
 immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g.,  
 25 lymph, serum, amniotic fluid, plasma, urine, synovial fluid and spinal fluid) or another  
 tissue or cell sample taken from an individual having such a disorder, relative to the  
 standard gene expression level, i.e., the expression level in healthy tissue or bodily  
 fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
 30 NO:129 as residues: Thr-17 to Trp-26, Pro-54 to Trp-61, Ala-65 to Arg-74, Pro-142 to  
 Leu-147, Pro-158 to Ala-165.

The tissue distribution in thymus and bone marrow cell lines indicates  
 polynucleotides and polypeptides corresponding to this gene are useful for the  
 diagnosis, treatment, and/or prevention of immune disorders involving stem cells.  
 35 Moreover, polynucleotides and polypeptides corresponding to this gene are useful for  
 the treatment and diagnosis of hematopoietic related disorders such as anemia,  
 pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are

important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the expression within fetal kidney indicates that this gene or gene product could be used in the treatment and/or detection of kidney diseases including renal failure, nephritis, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to Wilm's Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Moreover, the expression within fetal tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:15 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1002 of SEQ ID NO:15, b is an integer of 15 to 1016, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:15, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 6

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: HASAHASAHASGCGA (SEQ ID NO:306), QGVGVADGG

LERQRVDAGARLGHMGQPVAFSTRQLHLALPAPGTAGVTVPHPHAREGVV  
 GDLPLVPDAEDPTVGVPAGEGLLVLGHVVERAELILVRGLHQAEALARESEEMH  
 GSRHG (SEQ ID NO:307), EGGLERQRVDAGARLGHMGQPVAFS (SEQ ID  
 NO:308), LALPAPGTAGVTVPHPHAREGVVGDPLV (SEQ ID NO:309), PAEG  
 5 LLVLGHVVERAELILVRGLHQAEA (SEQ ID NO:310), HLFKFFYTIAFMQWF  
 TEFMELFLSVWELIKTXNLFCVCFSEHKPGQLVPAGPTSQLLCRALGRVH  
 LCSPTTRSQTPTQSWVTPQLLWRLGSGRLVAQVLQVGSFCGPRVGDAVLGEQT  
 FQPFDLL (SEQ ID NO:311), AFMQWTFEFMELFLSVWELIKTXNLFCVC (SEQ  
 ID NO:312), and/or RSQTPTQSWVTPQLLWRLGSGRLVAQ (SEQ ID NO:313).

10 Polynucleotides encoding these polypeptides are also encompassed by the invention.  
 The gene encoding the disclosed cDNA is believed to reside on chromosome 16.  
 Accordingly, polynucleotides related to this invention are useful as a marker in linkage  
 analysis for chromosome 16.

This gene is expressed primarily in human infant brain, and to a lesser extent, in  
 15 adult brain and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, neural, developmental, or pulmonary disorders. Similarly, polypeptides  
 20 and antibodies directed to these polypeptides are useful in providing immunological  
 probes for differential identification of the tissue(s) or cell type(s). For a number of  
 disorders of the above tissues or cells, particularly of the central nervous system  
 (CNS), expression of this gene at significantly higher or lower levels may be routinely  
 detected in certain tissues or cell types (e.g., neural, developmental, pulmonary, and  
 25 cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine,  
 amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken  
 from an individual having such a disorder, relative to the standard gene expression  
 level, i.e., the expression level in healthy tissue or bodily fluid from an individual not  
 having the disorder.

30 Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO:130 as residues: Ser-47 to Pro-57, Ser-77 to Glu-82, Thr-90 to Trp-98, Arg-124 to  
 Lys-137, Ala-183 to Glu-192, Lys-220 to Gln-229, Asn-244 to Arg-258, Thr-271 to  
 Asn-278, Glu-285 to Gly-297.

The tissue distribution in infant and adult brain indicates polynucleotides and  
 35 polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or  
 prevention of diseases of the CNS, such as mental retardation, schizophrenia,  
 Alzheimer's disease, paranoia, depression, and mania. Moreover, polynucleotides and

polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, dementia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The protein product of this gene may also show utility in the detection, treatment, and/or prevention of a variety of pulmonary disorders, particularly those related to disorders of the mucosal or endothelial tissues. The expression within fetal tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:16 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1225 of SEQ ID NO:16, b is an integer of 15 to 1239, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:16, and where b is greater than or equal to a + 14.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 7



The gene encoding the disclosed cDNA is believed to reside on chromosome 17. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 17. In specific embodiments, polypeptides of the invention

- 5 comprise the following amino acid sequence:  
 GAWGVEVVAVGSKAGCLVYQLCDLKQITFFFRASVCLSV (SEQ ID NO:314),  
 PASLGSSWGQKLRRGGTRKSFQELSPSSAPPAQLPQPPASTWLSSWPRPPCW  
 PPMCSWALGXCFPATGQWLPTSCCLWWCPDAGGRQKHFRSRWXTSWETW  
 QPYLTGLISSVLRAAXRPSYLQRFRLSXQXXLCCAFVIALGGGCFLLTALYLER  
 10 DETRAWQXVTGTPDSNDVDSNDLERQGLLSGXGASTEEL (SEQ ID NO:315), L  
 RGGTRKSFQELSPSSAPPAQLPQPP (SEQ ID NO:316), ATGQWLPTSCCLW  
 WCPDAGGRQKHFRSR (SEQ ID NO:317), GGCFLLTALYLERDETRAWQXV  
 (SEQ ID NO:318), APHLRLQPACHSPLPLPGSRPGPDHPAGLLCVPGPWG  
 ASVLQLGSGCRHPAVCGGAQMPGDGRSTSDHGGXHPGXPGSPISQDLSLVSC  
 15 GPXALTPICSASAAXXXXXCAAPLSSPWGAAASC (SEQ ID NO:319),  
 PACHSPLPLPGSRPGPDHPAGLLCV (SEQ ID NO:320), and/or  
 SGRHPAVCGGAQMPGDGRSTSDHGG (SEQ ID NO:321). Polynucleotides  
 encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in pineal gland and thymus.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, immune, endocrine, emotional or behavior disorders, in addition to  
 autoimmune diseases. Similarly, polypeptides and antibodies directed to these  
 25 polypeptides are useful in providing immunological probes for differential identification  
 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
 particularly of the immune system, expression of this gene at significantly higher or  
 lower levels may be routinely detected in certain tissues or cell types (e.g., immune,  
 endocrine, hematopoietic, neural, and cancerous and wounded tissues) or bodily fluids  
 30 (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or  
 cell sample taken from an individual having such a disorder, relative to the standard  
 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 individual not having the disorder.

- Preferred epitopes include those comprising a sequence shown in SEQ ID  
 35 NO:131 as residues: Asp-18 to Gln-27, Arg-44 to Asn-49.

The tissue distribution in thymus indicates polynucleotides and polypeptides  
 corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of

immune and autoimmune diseases, such as lupus, neutropenia, transplant rejection, and inflammatory diseases. Moreover, the expression within pineal gland indicates the protein product of this gene may be useful in disorders associated with biological clock aberrations, emotional distress, lethargy, or metabolic conditions. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:17 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1391 of SEQ ID NO:17, b is an integer of 15 to 1405, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:17, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 8

20

When tested against K562 cell lines, supernatants removed from cells containing this gene activated the ISRE (interferon-sensitive responsive element ) promoter element. Thus, it is likely that this gene activates leukemia cells, or more generally, immune or hematopoietic cells, through the JAK-STAT signal transduction pathway. ISRE is a promoter element found upstream in many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

GLKVMEICSLTFLEATNLQSRCQQAMLPLKALRKNPFLLLPSFDGCCQSLA  
FPGLWLQHSNLCLNHHMTFLVYLLCVSVFKYFFPFSCITYTSHWI (SEQ ID  
NO:322), ICSLTFLEATNLQSRCQQAMLP (SEQ ID NO:323), and/or GLWLQHS  
NLCLNHHMTFLVYLLCVSV (SEQ ID NO:324). Polynucleotides encoding these  
polypeptides are also encompassed by the invention.

This gene is expressed primarily in IL-1 and LPS induced neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, particularly inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:132 as residues: Ser-45 to His-50, His-52 to Ile-57, Lys-67 to His-81.

The tissue distribution in neutrophils, combined with the detected ISRE biological activity indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of inflammatory disorders, such as psoriasis, inflammatory bowel disease, rheumatoid arthritis, and sepsis. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:18 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1520 of SEQ ID NO:18, b is an integer of 15 to 1534, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:18, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: PFLLPPKRRGLLYHLIQKSTLGLVWVFREHLDSRSQ (SEQ ID NO:325), RGXPSWPMHTLVYAQHSTTHTPLIQPQWTQVIDQPPGITHQFCVR XCXCPTLESCVQECVTRSRXKPTTGVPGPQRLA (SEQ ID NO:326), TPLIQPQW TQVIDQPPGITHQFCV (SEQ ID NO:327), ALGPSQTCDLDVWLVAKPSFFRGPQ GIHYFSLWRRKPLSHWVSIWQLQGQETMPAMLRSRPAGQATVATGPPRGSPS PQDLPSYHRKQVESSHRHSWEPASQSQ (SEQ ID NO:328), CDLDVWLVAKPSF FRGPQGIHYFSLWRR (SEQ ID NO:329), and/or AGQATVATGPPRGSPSPQDLP SYHRKQV (SEQ ID NO:330). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in synovial fibroblasts, and to a lesser extent, in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skeletal or vascular disorders, particularly arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., skeletal, vascular, endothelial, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in synovial fibroblasts indicates polynucleotides and polypeptides corresponding to this gene are useful for treatment of arthritis. Moreover, polynucleotides and polypeptides corresponding to this gene are useful in the detection

and treatment of disorders and conditions afflicting the skeletal system, in particular osteoporosis, bone cancer, as well as, disorders afflicting connective tissues (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation), such as in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). The protein product of this gene may also be useful for the detection, treatment, and/or prevention of a variety of vascular disorders which include, but are not limited to, atherosclerosis, embolism, stroke, microvascular disease, or aneurysm. The protein may also be useful in the treatment of integumentary disorders, particularly those related to aberrations in the extracellular matrix or lamina. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:19 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1219 of SEQ ID NO:19, b is an integer of 15 to 1233, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:19, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 10

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

XGDTXTQNSRHDTPLIDYYRESCTLQYRPEFPGRPTRPRGSCPQYPGPAIPRT  
 SWALGEGDAAPRGAHH PRRADVPLG (SEQ ID NO:331), YRESCTLQYRPEFPG  
 RPTRPRGSCPQYPGP (SEQ ID NO:332), GKLYAAVPSGIPGSTHASARLMPPVS  
 RSSYSEDIVGSRRRRRSSSGSPSPQSRCSSWDGCSRSHSRGREGXRPPWSEL  
 DVGALYPFSRSGSRGRLPRFRNYAFASSWTSYSGYRYHRALLCRRTAVSGR  
 LREGREPSAEEAEGEREDWGIGSA (SEQ ID NO:333), SGIPGSTHASARLMPP

VSRSSYS (SEQ ID NO:334), GCSRSHSRGREGXRPPWSELVDVGALYPFS (SEQ ID NO:335), TAVSGRLREGREPSAEEAEGEREDW (SEQ ID NO:336), RIRKAA VQIPTRKNIGXRRPVVQETRKKERISRLKESIGNILIVTVTQSLTQILTLMMI KRELKPRRKRRKRNTKQXKRRIRKPKKNPVTQAVKTQKRTCQKLPGMEQ  
 5 PNVADTMDLIGPEAPINTYLFKMKNL (SEQ ID NO:337), TRKKERISRLKESI GNILIVTVTQSLTQ (SEQ ID NO:338), and/or VKTQKRTCQKLPGMEQPNVA DTMDLIGP (SEQ ID NO:339). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are  
 10 useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in retina.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
 15 not limited to, visual disorders, particularly retinopathy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the ocular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell  
 20 types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, vitreous humor, aqueous humor, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution in retina indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of diseases of the retina, for example, diabetic retinopathy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:20 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

35 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1076 of SEQ ID NO:20, b is an integer of 15

to 1090, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:20, and where b is greater than or equal to a + 14.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 11

When tested against NIH3T3 cell lines, supernatants removed from cells containing this gene activated the EGR1 (early growth response gene 1) promoter element. Thus, it is likely that this gene activates fibroblast cells, or more generally, other cells of the integumentary system, through the EGR1 signal transduction pathway. EGR1 is a separate signal transduction pathway from Jak-STAT. Genes containing the EGR1 promoter are induced in various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

LPFTLKPKMVKIPFSSRLINNNLQYIDCILSLKRCEEILLMWHGLLLCLASVFLE  
 LRGDRPPLLASLLEPHKMPLHSSSL (SEQ ID NO:340), LKPKMVKIPFSSRLIN  
 NNLQYIDC (SEQ ID NO:341), SLKRCEEILLMWHGLLLCLASVF (SEQ ID  
 NO:342), LRGDRPPLLASLLEPHKMPLH (SEQ ID NO:343), LQMHTGSGFGKGK  
 SCEVAFYVAQAEPGEGAYLHGAQETQKQGIEADHATLRGSPHSVSKTKYNLY  
 IANYYLLAWRKMES (SEQ ID NO:344), CEVAFYVAQAEPGEGAYLH (SEQ ID  
 NO:345), and/or ATLRGSPHSVSKTKYNLYIANY (SEQ ID NO:346).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

25 This gene is expressed primarily in human ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive disorders, particularly cancers, such as ovarian cancer.

30 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual  
 35 having such a disorder, relative to the standard gene expression level, i.e., the

expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in human ovarian cancer tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of ovarian cancer. Moreover, the expression within cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:21 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 668 of SEQ ID NO:21, b is an integer of 15 to 682, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:21, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 12

Contact of cells with supernatant expressing the product of this gene increases the permeability of the plasma membrane of myeloid leukemia cells to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of immune or hematopoietic cells, in addition to other cells or cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating myeloid cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:



- LSASLLDRYPASESNNYIFNFVLYMLHFLAGTLFSLFPDFELSPRSATLFPDLR  
TVQLLSSRPHL (SEQ ID NO:347), LLDRYPASESNNYIFNFVLYMLH (SEQ ID  
NO:348), FPDFELSPRSATLFPDLRTV (SEQ ID NO:349), NGGFYDVSFKQAG  
LIEFLCHYFYPMMAHVICGSRFTIVRTIPVHYVGEYFIKSSIWILYRINERTATKK  
5 AASDFQKNFRCFLDAF (SEQ ID NO:350), KQAGLIEFLCHYFYPMMAH (SEQ ID  
NO:351), and/or YFIKSSIWILYRINERTATKKAA (SEQ ID NO:352).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in anergic T-cells.

- Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, immune or hematopoietic disorders, particularly immunodeficiencies and  
inflammatory disorders. Similarly, polypeptides and antibodies directed to these  
polypeptides are useful in providing immunological probes for differential identification  
15 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
particularly of the immune system, expression of this gene at significantly higher or  
lower levels may be routinely detected in certain tissues or cell types (e.g., immune,  
hematopoietic cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum,  
plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken  
20 from an individual having such a disorder, relative to the standard gene expression  
level, i.e., the expression level in healthy tissue or bodily fluid from an individual not  
having the disorder.

- The tissue distribution in anergic T-cells, combined with the detected calcium  
flux biological activity, indicates polynucleotides and polypeptides corresponding to  
25 this gene are useful for the diagnosis and treatment of immune disorders, particularly  
those involving anergic T-cells. Moreover, the secreted protein can also be used to  
determine biological activity, to raise antibodies, as tissue markers, to isolate cognate  
ligands or receptors, to identify agents that modulate their interactions, and as  
nutritional supplements. It may also have a very wide range of biological activities.  
30 Typical of these are cytokine, cell proliferation/differentiation modulating activity or  
induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for  
treating human immunodeficiency virus infection, cancer, autoimmune diseases and  
allergy); regulation of hematopoiesis (e.g. for treating anaemia or as adjunct to  
chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or  
35 nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for  
control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections,  
tumors); hemostatic or thrombolytic activity (e.g. for treating haemophilia, cardiac

infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behaviour. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:22 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 756 of SEQ ID NO:22, b is an integer of 15 to 770, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:22, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 13

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: SPRXGRXFXTSRKQISGFLEFD (SEQ ID NO:353). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, particularly leukemia or multiple myeloma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the

expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in bone marrow tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of some types of leukemia, and other disorders involving bone marrow tissues or cells. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:23 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 551 of SEQ ID NO:23, b is an integer of 15 to 565, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:23, and where b is greater than or equal to a + 14.

### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 14

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: MKHAAFGLIPLVKEIYRYLKI KSKLLIGSGKCQLQPEWL QTSLINSSLLMDWLTPY (SEQ ID NO:354), IYRYLKI KSKLLIGSGKCQLQPE WLQTS (SEQ ID NO:355), QLGLPWDQSKGPRKNGLSMCGSVYSTIWSLIA SRREETIRVIVLYIQSPNINTRHISKRGLNKALTNP (SEQ ID NO:356), SKGPR

KNGLSMCGSVYSTIWS (SEQ ID NO:357), and/or QSPNINTRHISKRGLNK (SEQ ID NO:358). Polynucleotides encoding these polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in adult retina, and to a lesser extent, in 12 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, visual or developmental disorders, particularly retinopathy. Similarly,  
10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the ocular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., visual, developmental cancerous and wounded  
15 tissues) or bodily fluids (e.g., lymph, serum, amniotic fluid, aqueous humor, vitreous humor, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution in human retina indicates polynucleotides and polypeptides corresponding to this gene are useful for treatment of retinopathy, and other disorders involving the visual system. Moreover, the expression within embryonic tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other  
25 proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:24 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.  
35 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1342 of SEQ ID NO:24, b is an integer of 15

to 1356, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:24, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 15

The translation product of this gene has homology to the conserved human non-differentiated blood cell tyrosine kinase receptor fragment (See Genbank Accession No. R76466) which is thought to be important in signalling essential cellular pathways. In  
 10 specific embodiments, polypeptides of the invention comprise the following amino acid sequence: HPQTSAGGFPLHQGLPTVS (SEQ ID NO:359), PSWFPELSPWPLKTL KKRRQMRLRRRGRLCRLSPATTTTADTCRCPARSYRGSGRRPACAQDSPAPPS RPTRRAWEKALRPKRAAQWSTGVPPSPRSSTTGCCFGTAAXCAEGARR (SEQ ID NO:360), TTTADTCRCPARSYRGSGRRPA (SEQ ID NO:361), and/or  
 15 PSRPTRRAWEKALRPKRAAQWST (SEQ ID NO:362). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in human fetal epithelium.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental, integumentary, immune, or hematopoietic disorders, particularly skin cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
 25 particularly of the integumental system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., developmental, integumentary, immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having  
 30 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:139 as residues: Gln-26 to Ala-39, Cys-48 to His-55.

The tissue distribution in human fetal epithelium, combined with the homology  
 35 to a conserved tyrosine kinase receptor, indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of skin cancer, or other disorders related to the integument, particularly proliferative

conditions. Similarly, polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders including congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders such as arthritis, trauma, tendonitis, chondromalacia and inflammation, autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:25 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 603 of SEQ ID NO:25, b is an integer of 15 to 617, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:25, and where b is greater than or equal to a + 14.

## 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 16

When tested against PC12 cell lines, supernatants removed from cells containing this gene activated the EGR1 (early growth response gene 1) pathway. Thus, it is likely that this gene activates sensory neuron cells, or generally other cells or cell types, particularly immune cells, through the EGR1 signal transduction pathway. EGR1 is a separate signal transduction pathway from Jak-STAT, genes containing the EGR1 promoter are induced in various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

ARGVLNLRNRFECFSIIETV (SEQ ID NO:363), IGQLVMKSICHFQRLLSVAI  
10 DFASQFLKNYIFSSTHSSKAGFSVVCSLPKWLYTDGMEMVLKITHKLSF (SEQ  
ID NO:364), and/or QRLLSVAIDFASQFLKNYIFSSTH (SEQ ID NO:365).

Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 8. Accordingly, polynucleotides related to this invention are useful as a marker in  
15 linkage analysis for chromosome 8.

This gene is expressed primarily in fetal liver, and to a lesser extent, in resting T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune, hematopoietic, or hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of  
25 this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, hepatic, metabolic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, amniotic fluid, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,  
30 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal liver and resting T-cells, combined with the detected EGR1 biological activity indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of  
35 immune disorders involving T-cells, and more generally, hematopoietic conditions. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia,

pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Additionally, expression within fetal tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:26 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 634 of SEQ ID NO:26, b is an integer of 15 to 648, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:26, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 17

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LMKTASRMLLLE (SEQ ID NO:366). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in CD34-positive T cells from cord blood, and to a lesser extent, in anergic T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a



biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, particularly in immune system maturation and hematopoietic development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:141 as residues: Ile-46 to Tyr-56.

The tissue distribution in CD34-positive T cells and anergic T cells, indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of diseases involving hematopoietic development and stem cell maturation, including protection of stem cells from chemotherapy, immunosuppression during transplant rejection, and neutropenia. Moreover, this gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:27 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1374 of SEQ ID NO:27, b is an integer of 15 to 1388, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:27, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 18

When tested against U937 and Jurket cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) pathway. Thus, it is likely that this gene activates myeloid and T-cells, or more generally cells of immune or hematopoietic origin, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: ATXWDXPGCRNSARGERLHVGDAPW (SEQ ID NO:367), ARDER REVLKTLMLSTQRPQAFLPSQSWFVRLQKAGEGALKQENSLTIQNCLLCL PRVHRQRPTPPQPQRGNTEASVLQTSTEHLPRAAVLLVPNSCSPGXPTXLLSS (SEQ ID NO:368), ERREVLKTLMLSTQRPQAFLP (SEQ ID NO:369), GALKQEN SLTIQNCLLCLPRVHRQR (SEQ ID NO:370), and/or SVLQTSTEHLPRAAVLLVP NS (SEQ ID NO:371). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in activated human neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, such as neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic cancerous and wounded  
5 tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
10 NO:142 as residues: Val-25 to Gly-33.

The tissue distribution in activated neutrophils, combined with the detected GAS biological activity, indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of immune disorders involving neutrophils. Moreover, this gene product may be involved in the regulation of  
15 cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS,  
20 leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus  
25 erythematosis, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or  
30 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:28 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the  
35 scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of

a-b, where a is any integer between 1 to 602 of SEQ ID NO:28, b is an integer of 15 to 616, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:28, and where b is greater than or equal to a + 14.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 19

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: ALVISNPLL (SEQ ID NO:372), PYINTQMCVSSRNKFCISG  
 10 HQKYDSHGRETFRFEMHKARASSWKNILKIRSLKIISRGFEITNA (SEQ ID NO:373), KFCISGHQKYDSHGRETFRFEMHKARAS (SEQ ID NO:374), HTLLEI ANPLQAAVLGASSIHPSTHTSTHLMFMGLKWTELHHSPDSVQGAGAAEAAQTR HSLRPGRGRERHDCTLKNLTLFIIC (SEQ ID NO:375), NPLQAAVLGASSIHP SIHTSTH (SEQ ID NO:376), and/or SLRPGRGRERHDCTLKN (SEQ ID NO:377).  
 15 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
 20 not limited to, immune or hematopoietic disorders, such as neutropenia, inflammatory, or allergic conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or  
 25 lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 30 individual not having the disorder.

The tissue distribution in neutrophils indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of immune disorders involving neutrophils, or more generally, immune or hematopoietic disorders. Moreover, the expression of this gene product indicates a role in the  
 35 regulation of the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other

processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:29 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 814 of SEQ ID NO:29, b is an integer of 15 to 828, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:29, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in 7 week old early stage human.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fetal or developmental abnormalities, particularly congenital defects, including metabolic conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,

particularly of the developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in 7 week old early stage human tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of fetal abnormalities. Expression within embryonic tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein is useful in the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sachs disease, phenylketonuria, galactosemia, hyperlipidemias, porphyrias, leukemias, or Hurler's syndrome. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:30 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 567 of SEQ ID NO:30, b is an integer of 15 to 581, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:30, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 21

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

AENVHCTPAWETGRDSEDGKGREGMGRDRKGWDGTGLDGTGWEGKRERNV

PA (SEQ ID NO:378), GRDSEDGKGREGMGRDRKGWDGTGLD (SEQ ID NO:379), TSLGDLWDYNNSSH (SEQ ID NO:380), DRRIIRTREAAVAVSRERP LHSSLGNRERLRRWEGTGRDGKGQEGMGRDGTGWDGMGREERKKCPS (SEQ ID NO:381), RPLHSSLGNRERLRRWEGTGRDGKG (SEQ ID NO:382),  
 5 NQSWGPMGL (SEQ ID NO:383), GGGGCSEPRTSIALQPGKQGETPKMGRD GKGWEGTGRDGTGRDWMGRDGKGREKEMSQQ (SEQ ID NO:384), KQGE TPKMGRDGKGWEGTGRDGTG (SEQ ID NO:385), and/or PVLGTYGTITTPV TELTKGQEKEGGVETVLYE (SEQ ID NO:386). Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 This gene is expressed primarily in frontal cortex from a patient suffering from schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
 15 not limited to, neural disorders, such as Schizophrenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain  
 20 tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution in frontal cortex tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of some central nervous system disorders, for example, schizophrenia. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioural disorders, or  
 30 inflammatory conditions such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning  
 35 disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural

function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Moreover, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:31 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 775 of SEQ ID NO:31, b is an integer of 15 to 789, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID-NO:31, and where b is greater than or equal to a + 14.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 22

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: KIVFIDQKWSK (SEQ ID NO:387), CSLFWGILFLSRLRIH LFLSLKPCMCLRPIDILSHFLDIFVTSVLSELEKSSLKTTETFSFAVFLLLMMN (SEQ ID NO:388), LSRLRIHLFLSLKPCMCLRPIDILSH (SEQ ID NO:389), and/or VLSELEKSSLKTTETFSFAVFL (SEQ ID NO:390). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in thymus and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, particularly inflammation, or disorders related to immune cell maturation and/or activation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain



tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:146 as residues: Lys-38 to Leu-46.

The tissue distribution in thymus and neutrophils indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of inflammatory disorders, such as psoriasis, inflammatory bowel disease, rheumatoid arthritis, and sepsis. Moreover, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:32 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 870 of SEQ ID NO:32, b is an integer of 15 to

884, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:32, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 23

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: TLFYILH (SEQ ID NO:391). Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 This gene is expressed primarily in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases and disorders afflicting blood cells. Similarly, polypeptides and  
15 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or  
20 bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
25 NO:147 as residues: Pro-30 to Asn-36.

The tissue distribution in bone marrow indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of diseases afflicting the blood, including leukemia, neutropenia, anemia, and stem cell protection during chemotherapy. Moreover, polynucleotides and  
30 polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or  
35 chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the

expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

- 5 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:33 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.
- 10 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 852 of SEQ ID NO:33, b is an integer of 15 to 866, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:33, and where b is greater than or equal to a + 14.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in Merkel cells.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, integumentary disorders, particularly aberrations in mechanosensory function. Similarly, polypeptides and antibodies directed to these polypeptides are
- 25 useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in tissues involved in sensory function, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., integumentary, sensory, and cancerous and wounded tissues) or bodily fluids (e.g.,
- 30 lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 35 The tissue distribution in Merkel cells indicates polynucleotides and polypeptides corresponding to this gene are useful for Merkel cell dysfunctions, which may include aberrations in sensory function. Alternatively, polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or

- prevention of various skin disorders including congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's disease, mycosis fungoides, and Kaposi's sarcoma),
- 5 injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to
- 10 viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders such as arthritis, trauma, tendonitis, chondromalacia and inflammation, autoimmune disorders such as
- 15 rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets
- 20 for the above listed tissues.

- Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:34 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the
- 25 scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1680 of SEQ ID NO:34, b is an integer of 15 to 1694, where both a and b correspond to the positions of nucleotide residues shown
- 30 in SEQ ID NO:34, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 25

- 35 The translation product of this gene shares sequence homology with dihydropyridine receptor or nitrate transporter which are thought to be important in transport of small molecules across the cell membrane. In specific embodiments,

polypeptides of the invention comprise the following amino acid sequence:  
GTSFSVLSLIHDTG (SEQ ID NO:392). Polynucleotides encoding these polypeptides  
are also encompassed by the invention. The gene encoding the disclosed cDNA is  
believed to reside on chromosome 11. Accordingly, polynucleotides related to this  
5 invention are useful as a marker in linkage analysis for chromosome 11.

This gene is expressed primarily in kidney cortex and muscle tissue from a  
patient with multiple sclerosis, and to a lesser extent, in fetal liver/spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
10 biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, muscle, urogenital, or renal disorders, particularly musculodegenerative  
conditions such as multiple sclerosis, in addition to kidney or metabolic disorders and  
diseases. Similarly, polypeptides and antibodies directed to these polypeptides are  
useful in providing immunological probes for differential identification of the tissue(s)  
15 or cell type(s). For a number of disorders of the above tissues or cells, particularly of  
the multiple sclerosis and renal system, expression of this gene at significantly higher or  
lower levels may be routinely detected in certain tissues or cell types (e.g., muscle,  
urogenital, renal, hepatic, metabolic, immune, hematopoietic, and cancerous and  
wounded tissues) or bodily fluids (e.g., lymph, serum, bile, amniotic fluid, plasma,  
20 urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an  
individual having such a disorder, relative to the standard gene expression level, i.e.,  
the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
25 NO:149 as residues: Ala-66 to Leu-73.

The tissue distribution in kidney cortex and muscle tissue, combined with the  
homology to small molecule transporters indicates polynucleotides and polypeptides  
corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of  
disorders of renal functions and muscular diseases, including multiple sclerosis,  
30 muscular dystrophy, cardiomyopathy, fibroids, myomas, and rhabdomyosarcomas.  
The tissue distribution in kidney indicates that this gene or gene product could be used  
in the treatment and/or detection of kidney diseases including renal failure, nephritis,  
renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis,  
nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and  
35 kidney stones, in addition to Wilm's Tumor Disease, and congenital kidney  
abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome.  
Protein, as well as, antibodies directed against the protein may show utility as a tumor

marker and/or immunotherapy targets for the above listed tissues. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:35 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1201 of SEQ ID NO:35, b is an integer of 15 to 1215, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:35, and where b is greater than or equal to a + 14.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 26

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

35 VLISASTIGSRTSGAQGMEKMTIPTLA VGEPKTPEKSKCSLKQCFSSCNVHIDH  
LGLLLKCKF (SEQ ID NO:393), ASTIGSRTSGAQGMEKMTIPTLA (SEQ ID

NO:394), and/or GEPKTPEKSKCSLKQCFSSCNVHIDL (SEQ ID NO:395).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in kidney medulla.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, renal, urogenital, or more generally, disorders afflicting endothelial  
tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful  
10 in providing immunological probes for differential identification of the tissue(s) or cell  
type(s). For a number of disorders of the above tissues or cells, particularly of the renal  
system, expression of this gene at significantly higher or lower levels may be routinely  
detected in certain tissues or cell types (e.g., renal, urogenital, endothelial, and  
cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine,  
synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual  
15 having such a disorder, relative to the standard gene expression level, i.e., the  
expression level in healthy tissue or bodily fluid from an individual not having the  
disorder.

The tissue distribution in kidney medulla indicates polynucleotides and  
polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or  
20 prevention of renal disorders, including lesions, vascular diseases, hydronephrosis,  
and renal diseases associated with systemic disorders. Moreover, the gene or gene  
product could be used in the treatment and/or detection of kidney diseases including  
renal failure, nephritis, renal tubular acidosis, proteinuria, pyuria, edema,  
pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome,  
25 glomerulonephritis, hematuria, renal colic and kidney stones, in addition to Wilm's  
Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney,  
polycystic kidney, and Falconi's syndrome. The protein product can also be used for  
the treatment, detection, and/or prevention of various endothelial disorders, which  
include microvascular disease, embolism, aneurysm, stroke, or atherosclerosis.  
30 Protein, as well as, antibodies directed against the protein may show utility as a tumor  
marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available  
and accessible through sequence databases. Some of these sequences are related to SEQ  
ID NO:36 and may have been publicly available prior to conception of the present  
35 invention. Preferably, such related polynucleotides are specifically excluded from the  
scope of the present invention. To list every related sequence is cumbersome.  
Accordingly, preferably excluded from the present invention are one or more

polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1780 of SEQ ID NO:36, b is an integer of 15 to 1794, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:36, and where b is greater than or equal to a + 14.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

10 RIRSQDLAIMTTCFKKYEFSFFVLGFLRRWGATLCLGFTSFAIKFHPSSLCSEKE  
GKDFSGFALS IHGPERKKKEEGWARWLTPVVPVLWEAEVGGSPPEVSS (SEQ ID  
NO:396), TTCFKKYEFSFFVLGFLRRWGA (SEQ ID NO:397), SEKEGKDFSGF  
ALSIHGPERKKKEEGW (SEQ ID NO:398), MNECIAKPCMAAFSCSPSCCLPSR  
15 PGCSREQRCAFSCEPCHTVEHWVEPMGQQRQEHTQGSVLPSSHPSRGKATT  
VHSCCQEPWG (SEQ ID NO:399), FCSCPSCCLPSRPGCSREQRCAFSCEP (SEQ  
ID NO:400), GQRQEHTQGSVLPSSHPSRGKAT (SEQ ID NO:401), GVVNSCLL  
PLPPRLLATGMDCGGFASRRMGGRQHAALSVFLPLPLAHGLYPMFNCVAGLT  
GKGTSLLSGAARPAGEAAARAGTKGSHARFGNAFIHSFHSFIECLLNTYCV  
20 SSALTAVGIGDILKNKNDKSSCLCSC (SEQ ID NO:402), GMDCGGFASRRMG  
GRQHAALSVFLP (SEQ ID NO:403), LTGKGTSLLSGAARPAGEAAARAGT (SEQ  
ID NO:404), and/or LNTYCVPSALTAVGIGDILKN (SEQ ID NO:405).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal lung.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, pulmonary or developmental disorders and/or diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

30 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., pulmonary, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, pulmonary surfactant

35 or sputum, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression



level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal lung indicates polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of diseases related to pulmonary functions and infections. Moreover, the expression within fetal tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:37 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1160 of SEQ ID NO:37, b is an integer of 15 to 1174, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:37, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 28

This gene is expressed primarily in hepatocellular tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic or hepatic disorders or diseases, particularly hepatocellular tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., metabolic, hepatic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, bile, plasma, urine, synovial fluid and

spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution in hepatocellular tumors indicates polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hepatic disorders, particularly proliferative conditions such as hepatocellular tumors. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the  
10 differentiation of hepatocyte progenitor cells). In addition, the protein may play a role in the treatment, detection, and/or prevention of developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

15 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:38 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.  
20 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1073 of SEQ ID NO:38, b is an integer of 15 to 1087, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:38, and where b is greater than or equal to a + 14.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 29

30 The translation product of this gene has homology to a contains a helix-loop-helix motif from a *Caenorhabditis elegans* protein (See Genbank Accession No. g11326280) which is thought to function as a modulator of gene expression. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

35 TSLSQLWHFCHFVFPVKFCCGGCPVHCRMFSISGLYLLNASAPSLQLNDPKCL  
QT (SEQ ID NO:406), and/or WPVKFCCGGCPVHCRMFSISGLYLLNA (SEQ ID NO:407). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in normal breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, or endocrine disorders, particularly of the breast, such as breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer and metabolic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, endocrine, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, breast milk, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in breast indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of some types of breast cancer. The protein can also be used for the treatment, detection, and/or prevention of disorders related to ductile tissues or cell types, particularly secretory dysfunctions. The protein can also be used for the treatment of vascular disorders such as atherosclerosis, microvascular disease, embolism, stroke, or aneurysm. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:39 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 424 of SEQ ID NO:39, b is an integer of 15 to 438, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:39, and where b is greater than or equal to a + 14.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 30

When tested against K562 cell lines, supernatants removed from cells containing this gene activated the ISRE (interferon-sensitive responsive element) promoter element. Thus, it is likely that this gene activates leukemia cells, or potentially other cells or cell-types, through the JAK-STAT signal transduction pathway. ISRE is a promoter element found upstream in many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:  
SCRCWALGAGGGQRQWVGRS (SEQ ID NO:408), TGAQAPKMGARQKRPL  
QTRIKNSSKSTLWPPQWVRCGRWWTWPSRKKTSRPRRQLFTSTLSTSASALV  
WPVSWFSQEGH (SEQ ID NO:409), MGARQKRPLQTRIKNSSKSTLWPP (SEQ  
ID NO:410), and/or PRRQLFTSTLSTSASALVWPVSW (SEQ ID NO:411).  
Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in human testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive disorders, particularly abnormalities of the testes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, testicular, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, seminal fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:154 as residues: Leu-26 to Glu-52, Gln-71 to Lys-79.

The tissue distribution in testes, combined with the detected ISRE biological activity, indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of abnormalities of the testes, such as male infertility and proliferative conditions, and/or could be used as a male

contraceptive. The protein can also be used for the maintenance normal testicular function. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:40 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 720 of SEQ ID NO:40, b is an integer of 15 to 734, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:40, and where b is greater than or equal to a + 14.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 31

This gene is expressed primarily in colon, and to a lesser extent, in thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, gastrointestinal, immune, or hematopoietic disorders, particularly abnormalities of the colon, and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., gastrointestinal, immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in colon and thymus tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of abnormalities of the colon. The protein can also be used for treating inflammatory conditions, or potentially in modulating immune system activation in the treatment of gastrointestinal disorders. Protein, as well as, antibodies directed against

the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:41 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1332 of SEQ ID NO:41, b is an integer of 15 to 1346, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:41, and where b is greater than or equal to a + 14.

## 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 32

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: DGGGKEEGVSLKISLLCGPWLWLP (SEQ ID NO:412), HEMGELAICHTRVPFSLPSSAQGVQNLQGPIGHLAVCTPSSLTSHFHPQKREKW  
 20 STVNKRQRFLQFPAPLRNWIPQTPLSLSVSSGPLGSFTVFTLLSLCAWPWCCRD  
 CYKSCCPIPIFNLTAPLCVHTPEPSS (SEQ ID NO:413), SSAQGVQNLQGPIGHLAVCTPS (SEQ ID NO:414), VNKRQRFLQFPAPLRNWIPQTPLSLSVS (SEQ ID NO:415), CCRDCYKSCCPIPIFNLTAPLCV (SEQ ID NO:416), DLNVTNEGEGKE  
 25 VLGQGSTNNEKKKCQKATSNTEPRAREAKARHANMGTSRESPTWSLTAE  
 GLKAKSKMQGKATKGAASTMGSHNQGPHKREIFKHETPSSFPPPSQCQPE  
 LLPYKYWATLASGYVPSWLPSVDSYRINTAIKDKNGQDT (SEQ ID NO:417),  
 VLGQGSTNNEKKCQKA TSNTEPRA (SEQ ID NO:418), RESPTWSLTAE  
 GLKAKSKMQGKATKGAAS (SEQ ID NO:419), and/or GYVPSWLPSVDSYRI  
 NTAIKDK (SEQ ID NO:420). Polynucleotides encoding these polypeptides are also  
 30 encompassed by the invention.

This gene is expressed primarily in rhabdomyosarcoma, and to a lesser extent in heart and thymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
 35 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, muscle disorders, particularly rhabdomyosarcoma and other proliferative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., muscle, and cancerous and

5 wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID

10 NO:156 as residues: Gly-28 to Asp-33.

The tissue distribution in rhabdomyosarcoma tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of rhabdomyosarcoma. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of

15 various muscle disorders, such as muscular dystrophy, cardiomyopathy, fibroids, or myomas. The protein can also be used for the amelioration of proliferative conditions in other tissues, including modulation of the immune response to such tissues. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

20 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:42 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

25 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 984 of SEQ ID NO:42, b is an integer of 15 to 998, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:42, and where b is greater than or equal to a + 14.

30

### FEATURES OF PROTEIN ENCODED BY GENE NO: 33

In specific embodiments, polypeptides of the invention comprise the following

35 amino acid sequence: NSAEQSMILVT (SEQ ID NO:421), RXDRXPVPELPGYEPT  
RTDISSFKNIRYAFDFARDKDQRSLDIDTAKSMLALLLGRTWPLFSVFYQYLE  
QSKYRVMNKDQWYNVLEFSRTVHADLSNYDEDGAWPVLLDEFVEWQKVRQT

S (SEQ ID NO:422), PTRTDISSFKNIYRYAFDFARDKDQDSL (SEQ ID NO:423), SMLALLLGRTWPLFSVFYQYLE QSKYRVM (SEQ ID NO:424), FSRTVHADLSN YDEDGAWPVLLDEFVE (SEQ ID NO:425), IYRYAFDFAR (SEQ ID NO:426), KD QRSLDI (SEQ ID NO:427), NVLEFSRT (SEQ ID NO:428), and/or DLSNYDEDGA  
5 WPVLLDEFVEW (SEQ ID NO:429). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 11. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 11.

10 This gene is expressed primarily in aortic endothelium, and to a lesser extent, in cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endothelial disorders, particularly abnormalities of the vascular system  
15 and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., endothelial, vascular, and  
20 cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:157 as residues: Arg-22 to Lys-31.

The tissue distribution in aortic endothelium indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, detection, and/or prevention of abnormalities of the vascular system (i.e. embolism, atherosclerosis,  
30 aneurysm, stroke, microvascular disease, etc.) and cancers. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ  
35 ID NO:43 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.



Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 644 of SEQ ID NO:43, b is an integer of 15 to 658, where both a and b correspond to the positions of nucleotide residues shown in  
5 SEQ ID NO:43, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 34

10 Polypeptides of the invention do not comprise the polypeptide sequence shown as Genbank Accession W59652, which is hereby incorporated herein by reference. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LFRCPIGKAGTPAGXGPEFPGRPTRPVREKELTETFE (SEQ ID NO:430), FFVFPYPYFPRPLPPIPFPRFPWFRNFIPIPIESAPTTPLPSEK (SEQ ID  
15 NO:432), PWFRNFIPIPIESAPTTPLP (SEQ ID NO:433), and/or GKAGTPAGXGPEFPGRPTRPV (SEQ ID NO:431). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as  
20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, particularly Hodgkin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell  
25 type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual  
30 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:158 as residues: Ser-21 to Asp-35, Pro-47 to Pro-52, Pro-62 to Asn-67.

35 The tissue distribution in Hodgkin's lymphoma tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of Hodgkin's lymphoma. Moreover, polynucleotides and

polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic-related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex- vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:44 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 552 of SEQ ID NO:44, b is an integer of 15 to 566, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:44, and where b is greater than or equal to a + 14.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 35

When tested against U937 and K562 cell lines, supernatants removed from cells containing this gene activated both the GAS (gamma activating sequence), and the ISRE (interferon-sensitive responsive element ) promoter elements. Thus, it is likely that this gene activates pro-myeloid, leukemic, or more generally, other cells or cell-types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. ISRE is a promoter element found upstream in many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a

large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. The translation product of this gene was shown to have

5   homology to a conserved trypsin inhibitor which is thought to play an essential role in protein metabolism and regulation (See Genbank Accession No. pirlS35098IS35098). In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

FYPMTQGKESLPLLALQIFNTTFRPSFAFFSGHRTLFFGVRSPNPPKPRIFLIW  
 10   LIAVAL (SEQ ID NO:434), LLALQIFNTTFRPSFAFFSGHRTLFFGVRSP (SEQ ID NO:435), HLAQTVMMHPQKSFYQVKNTNHSRGAIEETQILEDRLGQIPLCLES QIWEA (SEQ ID NO:436), KNTNHSRGAIEET QILEDRLGQIPLCL (SEQ ID NO:437), QGCYRRDS NIGRQVRPDSIMLRKPDLGSIHYGSVLGNLNYCDLP QLYRNPSLGNSGMREMFSPFYNPVECHP (SEQ ID NO:438), PDSIMLRKPD  
 15   LGSITHYGSVLGN (SEQ ID NO:439), and/or YRNPSLGNSGMREMFSPFYNPV (SEQ ID NO:440). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in brain frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as

20   reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural disorders, particularly disorders of the central nervous system or endocrine system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the

25   tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system or endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample

30   taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain frontal cortex, combined with the detected GAS and ISRE biological activities indicates polynucleotides and polypeptides corresponding

35   to this gene are useful for diagnosis or treatment of disorders of the central nervous system, caused by trauma, inflammation, demyelination, neoplasia, and degenerative diseases. Additionally, the molecule may function as a neuropeptide or hormone.

Moreover, considering the homology to a trypsin inhibitor and its localization in the brain, indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:45 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1263 of SEQ ID NO:45, b is an integer of 15 to 1277, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:45, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 36

30

The translation product of this gene was found to have homology to a zinc finger protein from *Mus musculus* (See Genbank Accession No. gnlIPIDle225687) which is thought to be involved in the modulation of gene regulation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: NSARGLSGGHPFPWLSEGHPPF (SEQ ID NO:441), TDSDLTLGILLGI YTNHIWEMFLAASRINSPKLEPEKSVKRQINFSSKDVGCSLEVPKDGPPL SHGKEWIPLSHRKGWIPLSHMKGWPSLSHGKGWPP LSPRAEF (SEQ ID

NO:442), LGILLLLGIYTNHIWEMFLAA (SEQ ID NO:443), KSVKRQINFPSKDV  
 GCSLEVPKDGPP (SEQ ID NO:444), GKEWIPLSHRKGWIPLSHMKGWPSLSH  
 (SEQ ID NO:445), GWASTQPRERMDPAQPQERMDPSQPHERMALTQPWKMAP  
 TQPSCRI (SEQ ID NO:446), and/or PAQPQERMDPSQPHERMALTQPWK (SEQ ID  
 5 NO:447). Polynucleotides encoding these polypeptides are also encompassed by the  
 invention.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 10 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, immune or hematopoietic disorders. Similarly, polypeptides and  
 antibodies directed to these polypeptides are useful in providing immunological probes  
 for differential identification of the tissue(s) or cell type(s). For a number of disorders  
 of the above tissues or cells, particularly of the immune system, expression of this gene  
 15 at significantly higher or lower levels may be routinely detected in certain tissues or cell  
 types (e.g., immune, hematopoietic cancerous and wounded tissues) or bodily fluids  
 (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or  
 cell sample taken from an individual having such a disorder, relative to the standard  
 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 20 individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO:160 as residues: Ser-30 to Asp-39.

The tissue distribution in neutrophils indicates polynucleotides and polypeptides  
 corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of  
 25 immune disorders involving neutrophils, including neutropenia. The expression of this  
 gene product indicates a role in regulating the proliferation; survival; differentiation;  
 and/or activation of hematopoietic cell lineages, including blood stem cells. This gene  
 product may be involved in the regulation of cytokine production, antigen presentation,  
 or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by  
 30 boosting immune responses). Since the gene is expressed in cells of lymphoid origin,  
 the natural gene product may be involved in immune functions. Therefore it may be also  
 used as an agent for immunological disorders including arthritis, asthma,  
 immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis,  
 granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia,  
 35 neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune  
 reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-  
 host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue

injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The protein product of this gene can be used in the preparation of therapeutic compositions, for treating, preventing or delaying the recurrence of a tumour or neuronal disorders, e.g. genetic diseases or acquired degenerative encephalopathies such as Alzheimer's disease. Moreover, the protein is also useful in the induction or inhibition of cellular apoptosis resulting in inhibition of tumour cell growth, to suppress tumour formation, to induce G1 arrest of the cell cycle and to act as nuclear transcription factor. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:46 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 428 of SEQ ID NO:46, b is an integer of 15 to 442, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:46, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 37

When tested against U937 and K562 cell lines, supernatants removed from cells containing this gene activated both the GAS (gamma activating sequence), and the ISRE (interferon-sensitive responsive element ) promoter elements. Thus, it is likely that this gene activates pro-myeloid, leukemic, or more generally, other cells or cell-types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. ISRE is a promoter element found upstream in

many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. The protein product of this gene was found to have homology to the G-protein coupled receptor TM1 long consensus polypeptide (See Genbank Accession No. R50790) which indicates the protein is useful in the modulation of signalling events, cell-cycle regulation and/or transcriptional regulation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: IANGGGRPIKLNALYK IQNECKIVFTCIDF (SEQ ID NO:448), and/or MPCIK SKMNAKLFSLVLTLCMPISVLFGTCI (SEQ ID NO:449). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in duodenum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, gastrointestinal disorders, particularly abnormalities of the duodenum. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., gastrointestinal, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in duodenum, the homology to the TM1 g-protein coupled receptor consensus sequence, in addition to the detected GAS and ISRE biological activities, indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of the abnormalities of the duodenum, particularly proliferative conditions such as cancers. Moreover, the protein can be used in G-protein coupled receptor ligand binding assays. The assay can be used to identify fragments from GPR proteins (see Genseq Accession Nos. R48686-R48758 for examples) which retain biological activity such as binding a GPR ligand or

modulating GPR ligand binding to a GPR (see Genseq Accession Nos. R48759-  
R48758, R50569-R50807 and R89189-R89195 for examples of polypeptide  
fragments). The polypeptide fragments can be used in compositions for treating  
subjects suffering from a pathology related to a GPR abnormality e.g. a psychotic  
5 disorder such as schizophrenia. Protein, as well as, antibodies directed against the  
protein may show utility as a tumor marker and/or immunotherapy targets for the above  
listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available  
and accessible through sequence databases. Some of these sequences are related to SEQ  
10 ID NO:47 and may have been publicly available prior to conception of the present  
invention. Preferably, such related polynucleotides are specifically excluded from the  
scope of the present invention. To list every related sequence is cumbersome.  
Accordingly, preferably excluded from the present invention are one or more  
polynucleotides comprising a nucleotide sequence described by the general formula of  
15 a-b, where a is any integer between 1 to 876 of SEQ ID NO:47, b is an integer of 15 to  
890, where both a and b correspond to the positions of nucleotide residues shown in  
SEQ ID NO:47, and where b is greater than or equal to a + 14.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 38

The translation product of this gene shares sequence homology with a growth  
and transformation dependent protein (>gil207250), which is thought to be important in  
the regulation of cellular growth and proliferation. In specific embodiments,  
25 polypeptides of the invention comprise the following amino acid sequence:  
QVAMGSLSGRLRLAAGSCFRLCERDVSSSLRLTRSSDLKRINGFCTKPQESPG  
APSRTYNRVPLHKPTDWQKKILWSGRFKKEDIPETVSLEMLDAAKNK (SEQ  
ID NO:450), GLRLAAGSCFRLCERDVSSSLRLTR (SEQ ID NO:451), APSRTYNR  
VPLHKPTDWQKK (SEQ ID NO:452), IWSGRFKKEDIPETVSLEMLDA (SEQ ID  
30 NO:453), MDFAQNRKVPPELHPALTTECLYTNLRIGRKRSSYGQVASKRKM  
KSQRLSRWRCLMLQRTCE (SEQ ID NO:454), KVPPELHPALTTECLYTNLR  
(SEQ ID NO:455), KRSSYGQVASKRKMKSQRLSRWRCLM (SEQ ID NO:456),  
INGFCTKPQESP (SEQ ID NO:457), RVPLHKPTD (SEQ ID NO:458), WSGRFK  
KE (SEQ ID NO:459), EMLDAAKNK (SEQ ID NO:460), SYLMIALTV (SEQ ID  
35 NO:461), and/or MVIEGKKAA (SEQ ID NO:462). Polynucleotides encoding these  
polypeptides are also encompassed by the invention.



This gene is expressed primarily in ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, or endocrine disorders, particularly abnormalities of the ovary. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, endocrine, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:162 as residues: Lys-25 to Thr-33, Leu-39 to Glu-47.

The tissue distribution in ovary, combined with the homology to the growth and transformation dependent protein, indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of the abnormalities of the ovary such as ovarian cancer. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:48 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 723 of SEQ ID NO:48, b is an integer of 15 to 737, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:48, and where b is greater than or equal to a + 14.

35

## FEATURES OF PROTEIN ENCODED BY GENE NO: 39

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: RPGMRALGSCLSLALCSPQARPGPRTLDASTATLTPHF  
 5 SPCARFSPVGPSAVPFAATPLPLAGPHQP (SEQ ID NO:463), GSCLSLALCS  
 PQARPGPRT (SEQ ID NO:464), HFSPCARFSPVGPSAVPFAATPL (SEQ ID  
 NO:465), AIEERNKSRLTQQASEPTGSPRYLHEQHPSRSQMDCGSLTMXCPPP  
 RVRDDRTSARGVPRQAAPDIVGGRPSSRACVSPACAPSAAVFPY (SEQ ID  
 NO:466), LTQQASEPTGSPRYLHEQHPSRS (SEQ ID NO:467), and/or SARG  
 10 VPRQAAPDIVGGRPSSRACVS (SEQ ID NO:468). Polynucleotides encoding these  
 polypeptides are also encompassed by the invention.

This gene is expressed primarily in ovarian tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 15 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, reproductive or endocrine disorders, particularly ovarian tumors.  
 Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
 providing immunological probes for differential identification of the tissue(s) or cell  
 type(s). For a number of disorders of the above tissues or cells, particularly of the  
 20 female reproductive system, expression of this gene at significantly higher or lower  
 levels may be routinely detected in certain tissues or cell types (e.g., reproductive,  
 endocrine cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum,  
 plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken  
 from an individual having such a disorder, relative to the standard gene expression  
 25 level, i.e., the expression level in healthy tissue or bodily fluid from an individual not  
 having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO:163 as residues: Met-1 to Gly-6, Trp-23 to Arg-29, Ala-38 to Ser-45.

The tissue distribution in ovarian tumor tissue indicates polynucleotides and  
 30 polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or  
 prevention of reproductive disorders, particularly ovarian conditions, such as tumors.  
 Protein, as well as, antibodies directed against the protein may show utility as a tumor  
 marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available  
 35 and accessible through sequence databases. Some of these sequences are related to SEQ  
 ID NO:49 and may have been publicly available prior to conception of the present  
 invention. Preferably, such related polynucleotides are specifically excluded from the

- scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 557 of SEQ ID NO:49, b is an integer of 15 to 571, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:49, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 40

10

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: PRVRKTPHLSASGK (SEQ ID NO:469), YYYSMLKICHITI LETLSDRTPRKFAK KCYILYIKLSDSSVEKVAYTLLLLIPAAIEKK (SEQ ID NO:470), and/or TILETLSDRTPRKFAK KCYILYIKLSDSSVEK (SEQ ID NO:471).

- 15 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in endometrial stromal cells treated with estradiol.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive disorders, particularly cancer of the endometrium. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, endometrial, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:164 as residues: Met-1 to Ser-7.

- 35 The tissue distribution in endometrial stromal cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of diseases of the endometrium, particularly cancer or diseases caused by hormonal imbalances. Protein, as well as, antibodies directed against the protein may

show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:50 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 342 of SEQ ID NO:50, b is an integer of 15 to 356, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:50, and where b is greater than or equal to a + 14.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 41

The translation product of this gene shares sequence homology with the smaller hepatocellular oncoprotein which is thought to be important in protein synthesis leading to cellular transformation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: VHTKEIFRERSAGFPVK (SEQ ID NO:472), LEMGFQPTKEINARGSEPCQAQSTSLPKLPRWGSRPEAPQTPQGG LESRCCTPVSKQSLNLKADRFKALTLGRAQWLT PVIQALSELRWVDHLRSGV (SEQ ID NO:473), FQPT KEINARGSEPCQAQSTSLPK (SEQ ID NO:474), PKLPR WGSRPEAPQTPQGGLESRCCTP (SEQ ID NO:475), and/or RFKALTLGRAQWLT PVIQALSELRWVD (SEQ ID NO:476). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in human bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, urogenital disorders, particularly proliferative conditions, such as bladder tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bladder, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., urogenital, bladder, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal

fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:165 as residues: Pro-30 to Lys-38, Pro-45 to Ile-60, Leu-79 to Ser-96, His-98 to Gly-118.

The tissue distribution in bladder tumors indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis of carcinomas and preneoplastic or pathological conditions of bladder, or of the urogenital/renal system. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:51 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of  $a-b$ , where  $a$  is any integer between 1 to 899 of SEQ ID NO:51,  $b$  is an integer of 15 to 913, where both  $a$  and  $b$  correspond to the positions of nucleotide residues shown in SEQ ID NO:51, and where  $b$  is greater than or equal to  $a + 14$ .

## FEATURES OF PROTEIN ENCODED BY GENE NO: 42

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: RIPLQSDGSFLHEKSSQQRNRFPCPTLQCNPEVSFWFV VTDPSKNHTLPAVEVQSAIRMNKNRINNAFFLNDQTLEFLKIPSTLAPPMDP VPIWIIIFGVIFCIIIVAIALLLSGIWQRRRKNKEPSEVDDAEDKCENMITIENGIP SDPLDMKG GHINDAFMTEDERLTPL (SEQ ID NO:477), PCPTLQCNPEVSF WVVTDPSKNHT (SEQ ID NO:478), AIRMNKNRINNAFFLNDQTLEFL (SEQ ID NO:479), IWQRRRKNKEPSEVDDAEDKCENM (SEQ ID NO:480), PLDMKG GHINDAFMTEDER (SEQ ID NO:481), GSRTTALQRGVSLSSVMKASLICPP FMSRGSEGMPFSIVIMFSLSSASSTSDGSLFFLLRCQIPDKISSAIATMM MQNITPNIIQMGTDGSMGGASVEGIFKNSRVWSFRKKALLIRFLFILMADCTST A GRV (SEQ ID NO:482), VSLSSVMKASLICPPFMSRGSEGMPFS (SEQ ID

NO:483), and/or SMGGASVEGIFKNSRVWSFRKKAL (SEQ ID NO:484).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in kidney, and to a lesser extent, in gall bladder and testes.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases of the renal, urogenital, or reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing
- 10 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., renal, urogenital, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, seminal fluid, synovial
- 15 fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:166 as residues: Lys-60 to Ala-66, Thr-78 to Ser-83.

- 20 The tissue distribution in kidney indicates the protein product of this gene could be used in the treatment and/or detection of kidney diseases including renal failure, nephritis, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to Wilm's Tumor Disease, and congenital
- 25 kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Moreover, the tissue distribution in gall bladder indicates that the protein is useful for the treatment, detection, and/or prevention of various metabolic disorders. Alternatively, the expression within testes indicates that the protein is useful in normal testicular function. Therefore, this gene product may be useful in the treatment of male
- 30 infertility, and/or could be used as a male contraceptive. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

- Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ
- 35 ID NO:52 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1342 of SEQ ID NO:52, b is an integer of 15 to 1356, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:52, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 43

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GARGSQQDAPALQEAQVGRPERAQPARGR (SEQ ID NO:485), SERPGEGPARPGQDDQGPVPAVAGAGVGVHDPADHRVLGQRSAA HFYLHTSFSRPHTPGPDLTPGPDRTGSSRPTPMSTSFWTISHAGVKQSDLPKRKE TEQPPAPGEHGGGERERLRLVPARRPAQPRPGPAAGGAEERAAGLLRQLQP GLPHQGARIRRHQPQLGAEPDRGRPARGHLLRAQGGGLHQLEARDRAER KPAAPRCALPRPAAHPARARAQRQAPDLQQVLAPLREALPPPHEGQAQEVHQ VPLRARPLRAPDLRLPQQVRAGERGVLPQVRRHAAGVVRQPHQPARLGAR GLPRWPQGVLRQLHPVPAGPAHGEAGALQRALAAGVPPLPPVPDRLRFLG KLETLDAAQLLQLLQVDRQSASPRATGTGPPAAGRRTGSPRSPWPGG SSCINSTRPTLFSSATPSPKTSSETESFRVAFSRVPGT (SEQ ID NO:486), RPGQ DDQGPVPAVAGAGVGVHDP (SEQ ID NO:487), SRPHTGPPLTPGPDRT GSSR (SEQ ID NO:488), SHAGVKQSDLPKRKETEQPPAPGE (SEQ ID NO:489), RRPAQPRPGPAAGGAEERAAGLL (SEQ ID NO:490), RRHPQLGAEPDRGR PARGHLLL (SEQ ID NO:491), RDDRAERKPAAPRCALPRPAAHPAR (SEQ ID NO:492), RAPDLQQVLAPLREALPPPHEGQAQEV (SEQ ID NO:493), DLRLPQQ VRAGERGVLPQVRRHAAG (SEQ ID NO:494), QPARLGARGLPRWPQGVLR QLHPVPAG (SEQ ID NO:495), AGVPPLPPVPDRLRFLGKLETLD (SEQ ID NO:496), QLLQLLQVDRQSASPRATGTGPPAA (SEQ ID NO:497), NSTRPTLFSS ATPSPKTSSETESFR (SEQ ID NO:498), LGGKRTAGPPGVAAAAARRPRPE SPASPGIVVDLARVAEAVHLPPVLVEGRQLLRVRVQQVLDEVGEGHLEASA EGLARRGGQAGVVGVPQHGHGELAVELLVLQLELAAEGGDQAHEGVAHEE ELGVLELDLHEVAGELPVAAPELVEGQVRAGVVHVLARDAQRVAVGRTA VQQASQAQHDHHALPVGAGHLGHVAVDGPVPVVDQVAQLRVGDVVECALLG GEGQAGVGAEAPQHVPPLRLPALVWAAPGVARGPVVASHALLHAPPA QAAAPSPFWEGHSASRQHEKLSRNSSTSESAVSS LSCPARAWAAAAPCAA (SEQ ID NO:499), EAVHLPPVLVEGRQLLRVRVQQV (SEQ ID NO:500), GHLEA SAEGLARRGGQAGVVGVP (SEQ ID NO:501), QLELAAEGGDQAHEGVAHE

EELGVLLEL (SEQ ID NO:502), GELPVAAPELVEGQVRAGVVHVLARDA (SEQ ID NO:503), AVQQASAQHDHHPVGAGHLGHVA (SEQ ID NO:504), ALVW AAPGVARGPVVASHALLHA (SEQ ID NO:506), HDQVAQLRVGDVVECALLG GEGQAG (SEQ ID NO:505), PPAQAAAPSPFWEGHSASRQHEKLSRNS (SEQ ID NO:507), SRVTFPERRRSSRLRRGSMEESVRGYDWSPRDARRSPDQGRQQA  
 5 RRNVLRGFCANSSLAFTKERAFDDIPNSEL SHLIVDDRHGAIYCYVPKV ACTNWKRVMIVLSGSL LHRGAPYRDPLRIPREHVHNASAHLT FNKFWRRY GK LSRHLMKVKLK KYTKFLFVRDPFVRLISAFRSKFELENEEFYRKFAVPMLRVY ANHTSLPASAREAFRAGLKVSFANFIQYLLDPHTEKLAPFNEHW RQVYRLC  
 10 HPCQIDYDSWGSWRLWTRTPRSCCSYSRWTGSPLPELPEQDRQQLGGGLVR QD PPGLEAAAV (SEQ ID NO:508), RSPDQGRQAERRNVLRGFCANSSLA (SEQ ID NO:509), TKERAFDDIPNSEL SHLIVDDRHGAIYC (SEQ ID NO:510), FNKFWRRY GKLSRHLMKVKLK KY (SEQ ID NO:511), FVRLISAFRSKFELE NEEFYRKFA (SEQ ID NO:512), TSLPASAREAFRAGLKVSFANFIQYL (SEQ ID NO:513), and/or SYSRWTGSPLPELPEQDRQQLGGG (SEQ ID NO:514).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

The gene encoding the disclosed cDNA is believed to reside on chromosome 7.

Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 7.

20 It has been discovered that this gene is expressed primarily in PMA activated monocytic HL60 cells.

Therefore, nucleic acids of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of the following diseases and conditions: blood related disease such as  
 25 leukemia. Similarly, polypeptides and antibodies directed to those polypeptides are useful to provide immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be detected in certain tissues (e.g., immune, hematopoietic, and cancerous and wounded  
 30 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
 35 NO:167 as residues: Ala-29 to Thr-37, Pro-39 to Leu-63.

The tissue distribution in HL60 cells suggests the protein product of this clone is useful for the diagnosis, treatment, and/or prevention of blood related diseases such



as leukemia. Moreover, the protein product of this clone is useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:53 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence would be cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1533 of SEQ ID NO:53, b is an integer of 15 to 1547, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:53, and where b is greater than or equal to a + 14.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 44

When tested against fibroblast cell lines, supernatants removed from cells containing this gene activated the EGR1 (early growth response gene 1) promoter element. Thus, it is likely that this gene activates fibroblasts, or more generally, other cells or cell types, through the EGR1 signal transduction pathway. EGR1 is a separate signal transduction pathway from Jak-STAT, genes containing the EGR1 promoter are induced in various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation. The gene encoding the disclosed cDNA is believed to reside on chromosome 12. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 12. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: STGCSE (SEQ ID NO:515),

CLCLGCGLPELHSYLDPGPYLLVYPTLFWLCPSAVSPWAYTCYQLGLGPQWGA  
 AALSFTVDAAIRVWDVSTETCVPLPWFRGGGVNTNCSGPQTAAKSWLPLLQLSF  
 ESGRPRCGLVRGGLLYQGAVRLAA GAQMAADCCSLYWESH (SEQ ID  
 NO:516), YPTLFWLCPSAVSPWAYTCYQLGLGP (SEQ ID NO:517), DVSTETCV  
 5 LPWFRGGGVNTNCSGPQ (SEQ ID NO:518), LLYQGAVRLAA GAQMAADCCSL  
 (SEQ ID NO:519), NKRKTYLFLEVGMWGVGQNRWWPWVERVPRGRGWGCL  
 SKEGQVMNRASTPSRGFLGPPKHWAKTWKLIDKVQRDVGNSACGPAH  
 TEQGPFEGRWKVMSWGWAPGSPWIMPQGRSSNTGLFRVRKRRMTGLPS  
 CTLGFPFISTARRSPLGSQTME (SEQ ID NO:520), GVGQNRWWPWVERVPRG  
 10 RGWGCLSKEG (SEQ ID NO:521), AKTWKLIDKVQRDVGNSACGPAHTE  
 (SEQ ID NO:522), and/or WAPGSPWIMPQGRSSNTGLFRVRKRRMTGLPSC  
 TLGFPFIST (SEQ ID NO:523). Polynucleotides encoding these polypeptides are also  
 encompassed by the invention.

This gene is expressed primarily in fetal tissues such as fetal brain, fetal liver,  
 15 fetal kidney, and to a lesser extent, in T cells and macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, blood-related, immuno-related, neural-related, or developmental  
 20 disorders. Similarly, polypeptides and antibodies directed to these polypeptides are  
 useful in providing immunological probes for differential identification of the tissue(s)  
 or cell type(s). For a number of disorders of the above tissues or cells, particularly of  
 the hematopoiesis and immune system, expression of this gene at significantly higher or  
 lower levels may be routinely detected in certain tissues or cell types (e.g., neural,  
 25 immune, hematopoietic, urogenital, renal, hepatic, metabolic, developmental, and  
 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid,  
 bile, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from  
 an individual having such a disorder, relative to the standard gene expression level, i.e.,  
 the expression level in healthy tissue or bodily fluid from an individual not having the  
 30 disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO:168 as residues: Cys-126 to Thr-138, Glu-165 to Gly-172, Thr-189 to Leu-200,  
 Gly-222 to Gly-229, Pro-346 to Lys-354.

The tissue distribution in fetal liver indicates polynucleotides and polypeptides  
 35 corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of  
 blood related diseases, particularly immune or hematopoietic disorders. Alternatively,  
 the expression within fetal brain indicates polynucleotides and polypeptides

corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Alternatively, the expression within fetal kidney indicates the protein product of this gene could be used in the treatment and/or detection of kidney diseases including renal failure, nephritis, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to Wilm's Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Moreover, the expression within various fetal tissues, combined with the detected EGR1 biological activity, indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:54 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1324 of SEQ ID NO:54, b is an integer of 15 to 1338, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:54, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 45**

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: SSYQCPKVTFFKSSVDT (SEQ ID NO:524). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 15. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 15.

This gene is expressed primarily in glioblastoma, liver, fetal lung, and amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural, metabolic, or developmental disorders, particularly mental or neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, metabolic, developmental, pulmonary, hepatic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, pulmonart surfactant or sputum, bile, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:169 as residues: Pro-31 to Ala-37, Lys-62 to Asn-72.

The tissue distribution in glioblastoma and amygdala indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of central nervous system disorders. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder,

learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The protein may also be useful in the treatment, detection, and/or prevention of liver disorders, which include, but are not limited to hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:55 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2057 of SEQ ID NO:55, b is an integer of 15 to 2071, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:55, and where b is greater than or equal to a + 14.

25

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 46**

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: YIYSYLGFNFQINK (SEQ ID NO:525). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed in only T-cell helper II cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, particularly infectious diseases, inflammatory, or immunodeficiency conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:170 as residues: Pro-44 to Tyr-49.

The tissue distribution in T-helper cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment of infectious diseases. Moreover, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:56 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more

polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1885 of SEQ ID NO:56, b is an integer of 15 to 1899, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:56, and where b is greater than or equal to a + 14.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 47

The translation product of this gene has been shown to have homology to the human nuclear factor IV (See Genbank Accession No. gil35038), which is thought to play a role as a type 2 DNA helicase in DNA metabolism either during transcription, DNA repair, and/or during the cell-cycle. Moreover, the protein may play a role in chromosomal translocations. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: ARDLIL (SEQ ID NO:526), LTFYL QFLAPKDKPSGDTAAVFEEGGDVDDL VSTFNMHLVFCD (SEQ ID NO:527), and/or FLAPKDKPSGDTAAVFEEGGDVDDL (SEQ ID NO:528). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 2. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 2.

This gene is expressed primarily in activate T-cells, and to a lesser extent, in B-cells and monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or hematopoietic disorders, particularly leukemia, Grave's disease, rheumatoid arthritis and other autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:171 as residues: Gly-27 to Cys-35.

The tissue distribution in T-cells, B-cells, and monocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment or diagnosis of immune system diseases. Moreover, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's disease, scleroderma and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:57 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1529 of SEQ ID NO:57, b is an integer of 15 to 1543, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:57, and where b is greater than or equal to a + 14.

35



## FEATURES OF PROTEIN ENCODED BY GENE NO: 48

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: ARAGAKILFEGEF (SEQ ID NO:529), NFEIHSAPFMLFVA  
 5 CLLHSSCPRTARFLASPLSESNVIFYQNQYQFPCILCFIEFARLTSFKHLIHSQSH  
 LVRLQYEDFSVSSE AWDTELT (SEQ ID NO:530), FPFMLFVACLLHSSCPRTA  
 RFLASPL (SEQ ID NO:531), NVIFYQNQYQFPCILCFIEFARLTSF (SEQ ID  
 NO:532), and/or SQSHLVRLQYEDFSVSSE AWDTE (SEQ ID NO:533).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

10 The gene encoding the disclosed cDNA is believed to reside on chromosome 14.  
 Accordingly, polynucleotides related to this invention are useful as a marker in linkage  
 analysis for chromosome 14.

This gene is expressed primarily in fetal tissues such as fetal liver, fetal brain,  
 fetal lung and fetal spleen.

15 Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, developmental disorders and cancers. Similarly, polypeptides and  
 antibodies directed to these polypeptides are useful in providing immunological probes  
 20 for differential identification of the tissue(s) or cell type(s). For a number of disorders  
 of the above tissues or cells, particularly of the immune system and nervous system,  
 expression of this gene at significantly higher or lower levels may be routinely detected  
 in certain tissues or cell types (e.g., developmental, hepatic, immune, hemaopoietic,  
 neural, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum,  
 25 plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell  
 sample taken from an individual having such a disorder, relative to the standard gene  
 expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
 30 NO:172 as residues: Gly-37 to Asp-46, Ser-48 to Val-54.

The tissue distribution in fetal tissues indicates polynucleotides and polypeptides  
 corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of  
 developmental disorders and cancers. Moreover, the expression within embryonic  
 tissue and other cellular sources marked by proliferating cells indicates that this protein  
 35 may play a role in the regulation of cellular division, and may show utility in the  
 diagnosis and treatment of cancer and other proliferative disorders. The protein is also  
 useful in the treatment, detection, and/or prevention of immune, hematopoietic,

pulmonary, or metabolic diseases, disorders, and/or conditions. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:58 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1119 of SEQ ID NO:58, b is an integer of 15 to 1133, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:58, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene was found to have homology to a 35kd pulmonary surfactant protein, as well as, a GABA-like receptor (See Genbank Accession Nos. P70663, and gil540271, respectively), the latter of which is thought to be important in neuronal function. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: QKFLCASDGD (SEQ ID NO:534), AEVPLRVRRRHGRPHGPGGRQLALGIPALRSLPGCVPRHHGC SPGYGCLHRRILCLPLILLVYKQRQAASNRAQELVRMDSNIQGIENPGF EASPPAQGIPEAKVRHPLSYVAQRQPSESGRHLLSEPSTPLSPPGPDVFF PSLDPVPDSPNFVIXPXWGTVGCCGWVWGRCI (SEQ ID NO:535), GPGG RQLALGIPALRSLPGCVPRHHGC (SEQ ID NO:536), FEASPPAQGIPEAK VRHPLSYVAQR (SEQ ID NO:537), DMSLGMWQHQQWDMKMDTGPPSQAPD TGHGGETSPPWHALGSPVLPEAALLSDFLFPQWLWGQACLPTGHRHLPQLPP TSSF SEDLSTG (SEQ ID NO:538), PPSQAPDTGHGGETSPPWHALGS (SEQ ID NO:544), PVDRSSEKLLVGGSWGRWRWPVGRQAWPQSHCGTKRKSDRR AASGKTGEPSACHGGEVSPPCPVSGAWEG GPVSILSH (SEQ ID NO:539), PVDRSSEKLLVGGSWGRWRWPV (SEQ ID NO:540), TKRKSDRRAASG KTGEPSACHGGEV (SEQ ID NO:541). MTSKFGESGTGSRDGKKTSPGPG

GDRGVLGSESRCPDSEGCRWAT (SEQ ID NO:542), and/or SPGPGGDRGV LGSESRCPD (SEQ ID NO:543). Polynucleotides encoding these polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in hematopoiesis cells such as neutrophils, eosinophils and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, blood diseases and/or immune diseases. Similarly, polypeptides and  
10 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and  
15 wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
20 NO:173 as residues: Ser-44 to Ala-63, Pro-89 to Gly-98, Pro-129 to Trp-137.

The tissue distribution in neutrophils, eosinophils, and T cells indicates polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosis blood related diseases. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic  
25 related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be  
30 used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The homology to a pulmonary surfactant protein indicates that the protein is useful in enhancing or inhibiting the efficacy of the  
35 immune response across mucosal barriers, such as within the gastrointestinal tract, the sinuses, and the lungs. Protein, as well as, antibodies directed against the protein may

show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:59 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1476 of SEQ ID NO:59, b is an integer of 15 to 1490, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:59, and where b is greater than or equal to a + 14.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 50

When tested against U937 cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates promyeloid cells, or more generally, other cells of the immune or central nervous system, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: HEVQPSYLPSNSGLI (SEQ ID NO:545). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in the central nervous system, adult liver, adult heart, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural, cardiovascular, or metabolic conditions or disorders. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely  
5 detected in certain tissues or cell types (e.g., neural, cardiovascular, developmental, metabolic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
10 individual not having the disorder.

The tissue distribution in tissues of the CNS and infant brain, combined with the detected GAS biological activity indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of CNS disorders. Moreover, polynucleotides and polypeptides corresponding to this  
15 gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction,  
20 aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene  
25 product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The expression within fetal tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or  
30 apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available  
35 and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:60 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the

scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1322 of SEQ ID NO:60, b is an integer of 15 to 1336, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:60, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 51

10

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LRISVLCRETACNWSHHPLDSN (SEQ ID NO:546).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

The gene encoding the disclosed cDNA is believed to reside on chromosome 18.

15 Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 18.

This gene is expressed in whole brain, embryos, fetal liver and fetal spleen, and melanocytes.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural, immune, hematopoietic, or developmental disorders, particularly mental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, immune, hematopoietic, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:175 as residues: Pro-27 to Lys-42.

35 The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of mental or neurodegenerative disorders. Alternatively, the expression within fetal

liver/spleen indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include

5 bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors

10 of various blood lineages, and in the differentiation and/or proliferation of various cell types. The protein may also be useful for the treatment and/or detection of metabolic disorders, which include Tay-Sachs disease, phenylketonuria, galactosemia, hyperlipidemias, porphyrias, and Hurler's syndrome. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy

15 targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:61 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the

20 scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1691 of SEQ ID NO:61, b is an integer of 15 to 1705, where both a and b correspond to the positions of nucleotide residues shown

25 in SEQ ID NO:61, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 52

30 When tested against U937 and fibroblast cell lines, supernatants removed from cells containing this gene activated both the GAS (gamma activating sequence) and EGR1 (early growth response gene 1) promoter elements. Thus, it is likely that this gene activates promyeloid cells, fibroblasts, or more generally, immune or integumentary cells or cell-types, through the JAK-STAT and/or EGR1 signal

35 transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells.

Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. EGR1 is a separate signal transduction pathway from Jak-STAT, genes containing the EGR1 promoter are induced in various tissues and cell types upon  
5 activation, leading the cells to undergo differentiation and proliferation. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LTVTVRNPGSTHASGRPRRRSGVWARRGLVWQ (SEQ ID NO:547). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in endometrial stromal cells and fetal brain  
10 tissue, and to a lesser extent, in microvascular endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, neural, developmental, or vascular disorders, particularly  
15 vascular leak syndrome and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endothelium, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell  
20 types (e.g., reproductive, neural, developmental, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
25 NO:176 as residues: Pro-63 to Cys-72, Gly-88 to Cys-93.

The tissue distribution in endometrial stromal cells, infant brain, and microvascular endothelial cells, combined with the detected GAS and EGR1 biological activities, indicates polynucleotides and polypeptides corresponding to this gene are  
30 useful for the treatment of various vascular disorders, which include, but are not limited to vascular leak syndrome, microvascular disease, atherosclerosis, aneurysm, stroke, embolism and inflammation. Moreover, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and  
35 treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could



again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:62 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1017 of SEQ ID NO:62, b is an integer of 15 to 1031, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:62, and where b is greater than or equal to a + 14.

15

### FEATURES OF PROTEIN ENCODED BY GENE NO: 53

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of HUVEC cells to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both vascular endothelial cells, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating endothelial cells, or more generally, neural or immune cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell. This protein is homologous to members of the butyrophilin gene family which are thought to play a role in myelin sheath development, in addition to serving as a membrane-specific receptor for cytoplasmic vesicles to the apical plasma membrane. In specific embodiments, polypeptides of the invention comprise the sequence  
SAQFSVLGSPGPILAMVGEDADLPCHLFPTMSAETMELKW (SEQ ID NO:574). Polynucleotides encoding these polypeptides are also encompassed by the invention. In specific embodiments, polypeptides of the invention comprise the sequence  
TPCSAQFSVLGSPGPILAMVGEDADLPCHLFPTMSAET (SEQ ID NO:548),  
MELKWVSSSLRQVVNVYADGKEVEDRQSAPYRGRTSILRDGITAGKAALRIHN

VTASDSG (SEQ ID NO:549), LEVKG YEDGGIHLECRSTGWYPQPI (SEQ ID NO:550), MASSLAFLLLNFHVSLLLVQLLTPCSAQFSVLGSPGPILAMVGE DADLPCHLFTMSAETMELK WVSSSLRQVVNVYADG (SEQ ID NO:551), RHELSHNRKNGELLIDRLYSVGSDSPMGIPRDIIFTDGFPYWNPKVKT LKDRHF

5 WQSIDENGKFPGFPSA QLSCLPPLGPAAHSLSSVFCATLWAHPGHGG (SEQ ID NO:552), LLIDRLYSVGSDSPMGIPRDIIFT (SEQ ID NO:553), NPKVKT LKDRHFWQSIDENGKFPGF (SEQ ID NO:554), LGPAAHSLSSVFCATLWA HPGH (SEQ ID NO:555). RLQHWVLIFTLEVKG YEDGGIHLECRSTGWYPQ

10 QIQWSNAKGENIPAVEAPVVADGVGLYEVAASVIMRGGSGEGVSCIIRNSLL GLEKTASISIADPSSGAPSPGSQPWQGPCLSCCCFSPEPVTSCGDNRK (SEQ ID NO:556), GGIHLECRSTGWYPQPIQWSNAKG (SEQ ID NO:557), PQIQWS NAKGENIPAVEAPVVADGVGL (SEQ ID NO:558), NIPAVEAPVVADGVGL YEVAASVIMRG (SEQ ID NO:559), SGAPSPGSQPWQGPCLSCCCFSPEPVT (SEQ ID NO:560), SSSICDHERRLRGGCILHHQKFP RPKGDSQHFHRRP

15 FFRSAQPWIAALAGTLPI LLLLLAGASYFLWRQQKEITALSSEIESEQEMKE MG YAATEREISLRESLQEELKRKKIQYLTRGESSSDTNKSA (SEQ ID NO:561), KDSQHFHRRPFFRSAQPWIAALAGTLPI (SEQ ID NO:562), EIESEQEMKE MG YAATEREISLRESLQE (SEQ ID NO:563), VNNMIAFY SARDSYVYPHFSG EEMLQMLHLVK (SEQ ID NO:564), TPCSAQFSVLGSPGPILAMVGEDADLP

20 CHLFTMSAET (SEQ ID NO:565), KWVSSSLRQVVNVYADGKEVEDR (SEQ ID NO:566), RTSILRDGITAGKAALRIHNV T ASD (SEQ ID NO:567), CYFQDGDYF EKALVELKVAALGS (SEQ ID NO:568), GYEDGGIHLECRSTGWYPQPIQ (SEQ ID NO:569), NIPAVEAPVVADGVGLYEVAASV (SEQ ID NO:570), QQKEITALSS EIESEQEMKEM (SEQ ID NO:571), LRESLQEELKRKKIQYLTRGESS (SEQ ID

25 NO:572), and/or GEEMLQMLHLVK (SEQ ID NO:573). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in rhabdomyosarcoma, and to a lesser extent,

30 in T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, muscle, immune, or neural disorders, particularly rhabdomyosarcoma,

35 infectious diseases, or neurodegenerative conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., muscle, immune, neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:177 as residues: Ala-78 to Arg-94.

The tissue distribution in rhabdomyosarcoma, the detected calcium flux biological activity, combined with the homology to the butyrophilin gene family indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis or treatment of muscle disorders, which include, but are not limited to, muscular dystrophy, cardiomyopathy, fibroids, myomas, and/or rhabdomyosarcomas.

Moreover, the homology to the butyrophilin protein indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The protein may also show utility in the correction or amelioration of myelin sheath deficiencies in developing and mature neurons and neural-cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:63 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more

polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1575 of SEQ ID NO:63, b is an integer of 15 to 1589, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:63, and where b is greater than or equal to a + 14.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 54

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: PQGGLTLPSVWG (SEQ ID NO:575), GGPCHLWLLGPRRT QLPGRRASLPFRSQGELTQAFLLGLWKHQMPALTQEQQVRAERRREA VRMEI PGLFFASLANWGLLYRTSQDFISPYLCAAPSTPHPPLGGP (SEQ ID NO:576), GPRRTQLPGRRASLPFRSQGELT (SEQ ID NO:577), QMPALTQEQQVRAER RREA VRMEI (SEQ ID NO:578), and/or ANWGLLYRTSQDFISPYLCAAPSTP (SEQ ID NO:579). Polynucleotides encoding these polypeptides are also encompassed by the invention. The gene encoding the disclosed cDNA is believed to reside on chromosome 5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5.

This gene is expressed primarily in brain, and to a lesser extent, in testes tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural, endocrine, or reproductive disorders, particularly depression and infertility disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, endocrine, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, seminal fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:178 as residues: Thr-26 to Glu-33.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of depression and

other endocrine-related disorders. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, 5 Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in 10 feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The protein product may also be useful in the treatment, detection, and/or 15 prevention of a variety of reproductive disorders which include, but are not limited to, the treatment of male infertility, and/or could be used as a male contraceptive. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available 20 and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:64 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more 25 polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1074 of SEQ ID NO:64, b is an integer of 15 to 1088, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:64, and where b is greater than or equal to a + 14.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 55

The translation product of this gene was found to have homology to the conserved R166.2 protein from *Caenorhabditis elegans* (See Genbank Accession 35 No.gil949849), which is thought to play an important role in the regulation of cellular function and processes. In specific embodiments, polypeptides of the invention comprise the sequence: LSFKDKSTYIESSTKVDMAFRYLSWILFPLL (SEQ ID

NO:580), CYAVYSLLYLEHKGWYSWVLSM (SEQ ID NO:590), LLTFGFITMTPQ  
 LFINYKLKSV AHL PWRMLT (SEQ ID NO:581), TYKALNTFIDDLFAFVIKMP  
 VMYRIGCLRD (SEQ ID NO:582), DVVFFIYLYQRWIYRVDPTRVNEFGMSGED  
 (SEQ ID NO:583), VAGIFPRLSFKDKSTYIESSTKVYDDMAFRYLSWILFPLL  
 5 CYA (SEQ ID NO:584), PWVAGIFPRLSFKDKSTYIESSTKVYDD (SEQ ID  
 NO:586), AGEDSCHPVLSVQPDVHDLGWQESSPAYPSRTSPRISSPRPKC  
 MMIWHSGETCPGSSSR SWAAMPSTVFCTWSTRAGTPGCSACSTASC (SEQ ID  
 NO:587), LSVQPDVHDLGWQESSPAYPSRTSPRISSP (SEQ ID NO:588), GSSSR  
 SWAAMPSTVFCTWSTRAGTP (SEQ ID NO:589), and/or WAAMPSTVFCTWS  
 10 TRAGTP (SEQ ID NO:585). Polynucleotides encoding these polypeptides are also  
 encompassed by the invention. The gene encoding the disclosed cDNA is believed to  
 reside on chromosome 19. Accordingly, polynucleotides related to this invention are  
 useful as a marker in linkage analysis for chromosome 19.

This gene is expressed primarily in colon, smooth muscle and fetal bone.

15 Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 not limited to, gastrointestinal or vascular disorders, and abnormal muscular-skeletal  
 development, including proliferative conditions such as cancer. Similarly, polypeptides  
 20 and antibodies directed to these polypeptides are useful in providing immunological  
 probes for differential identification of the tissue(s) or cell type(s). For a number of  
 disorders of the above tissues or cells, particularly of the immune and muscular-skeletal  
 system, expression of this gene at significantly higher or lower levels may be routinely  
 detected in certain tissues or cell types (e.g., gastrointestinal, muscle, skeletal, vascular,  
 25 developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum,  
 plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell  
 sample taken from an individual having such a disorder, relative to the standard gene  
 expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 individual not having the disorder.

30 Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO:179 as residues: Ser-128 to Thr-133, Thr-158 to Thr-166, Leu-168 to Gly-175,  
 Ala-179 to Asp-196.

The tissue distribution in colon and fetal bone indicates polynucleotides and  
 polypeptides corresponding to this gene are useful for the treatment and diagnosis of  
 35 abnormal bone formation, and/or various proliferative conditions (e.g. tumors),  
 particularly of the gastrointestinal system. Moreover, the expression within smooth  
 muscle tissue indicates polynucleotides and polypeptides corresponding to this gene are

useful for the detection, treatment, and/or prevention of a variety of vascular disorders, which include, but are not limited to the following: embolism, atherosclerosis, microvascular disease, aneurysm, stroke, and vascular leak syndrome. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:65 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1242 of SEQ ID NO:65, b is an integer of 15 to 1256, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:65, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 56

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: LGEFLSSQCFLP (SEQ ID NO:591). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in brain frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural disorders, particularly neurological or neurodegenerative disorders and diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:180 as residues: Ala-122 to Gly-128.

The tissue distribution in brain frontal cortex indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of some neurological diseases such as depression. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:66 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1588 of SEQ ID NO:66, b is an integer of 15 to 1602, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:66, and where b is greater than or equal to a + 14.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 57

When tested against U937 cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates pro-myeloid cells, or more generally, immune or hematopoietic cells, through the JAK-STAT signal transduction pathway. GAS is a

35



promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:  
RSRRNRVAMGMWASLDALWE (SEQ ID NO:592), PRVRCQQRAEGGMGAG  
IGVGPSERTDIAVTPRGRSEGASVGVAPVHAEGAGGTGWPWGCGHRWTLCG  
RCR PRSVSSGPCCSFPGQCIFGRPS (SEQ ID NO:593), GGMGAGIGVGPSER  
TDIAVTPRGR (SEQ ID NO:594), GCGHRWTLCGRCR PRSVSSGPCCSFP (SEQ  
ID NO:595), and/or KKHGF NQQLGFFTWKYNKNKNLV (SEQ ID NO:596).  
Polynucleotides encoding these polypeptides are also encompassed by the invention.  
The gene encoding the disclosed cDNA is believed to reside on chromosome 1.  
Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in synovial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skeletal afflictions, particularly rheumatoid arthritis or autoimmune conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., skeletal, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:181 as residues: Gln-27 to Val-39, Glu-50 to Arg-56.

The restricted tissue distribution in synovium, combined with the detected GAS biological activity, indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of rheumatoid arthritis since synovial fibroblasts are associated with the synovium and cartilage. Moreover, polynucleotides and polypeptides corresponding to this gene are useful in the detection and treatment of disorders and conditions afflicting the skeletal system, in particular osteoporosis, bone

cancer, as well as, disorders afflicting connective tissues (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation), such as in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). The protein may also be useful in the modulation of the immune response to regions of inflammation, or in inhibiting or ameliorating autoimmune responses. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:67 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 924 of SEQ ID NO:67, b is an integer of 15 to 938, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:67, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 58

When tested against U937 cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates pro-myeloid cells, or more generally immune or hematopoietic cells, in addition to other cells or cell-types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: PKLLPCSPAEGHTSLGPLLPF (SEQ ID NO:597), ASLELXPS KSQLSTEWGFTWIVGLGMSPSTALWTECTCTPFLVLLSHASGHFFWLSPLAS

LVIPPVTDRK (SEQ ID NO:598), WGFTWIVGLGMSPSTALWTECTCTPFLVL  
 LSH (SEQ ID NO:599), VAVGVCREDMGITDRSKMSPDVGIWAIYWSAAGY  
 WPLIGFPGTPTQQEPALHRVGVYLDRTGNVSFYSAVDGVHLHTFSCS  
 SVSRLRPFFLVESISIFSHSTSD (SEQ ID NO:600), ITDRSKMSPDVGIWAIYW  
 5 SAAGYWPLI (SEQ ID NO:601), and/or RGTGNVSFYSAVDGVHLHTFSCSSV  
 SRLRP (SEQ ID NO:602). Polynucleotides encoding these polypeptides are also  
 encompassed by the invention. The gene encoding the disclosed cDNA is believed to  
 reside on chromosome 7. Accordingly, polynucleotides related to this invention are  
 useful as a marker in linkage analysis for chromosome 7.

10 This gene is expressed primarily in fetal tissues, and to a lesser extent, in liver  
 cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
 reagents for differential identification of the tissue(s) or cell type(s) present in a  
 biological sample and for diagnosis of diseases and conditions which include, but are  
 15 not limited to, developmental or liver diseases, such as hepatocellular carcinoma.  
 Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
 providing immunological probes for differential identification of the tissue(s) or cell  
 type(s). For a number of disorders of the above tissues or cells, particularly of the  
 immune and hepatic systems, expression of this gene at significantly higher or lower  
 20 levels may be routinely detected in certain tissues or cell types (e.g., developmental,  
 hepatic, metabolic, and cancerous and wounded tissues) or bodily fluids (e.g., serum,  
 plasma, bile, breast milk, urine, synovial fluid and spinal fluid) or another tissue or cell  
 sample taken from an individual having such a disorder, relative to the standard gene  
 expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
 25 individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
 NO:182 as residues: Pro-30 to Gln-37, Arg-39 to Ser-45, Arg-74 to Arg-85.

The tissue distribution in liver, combined with the detected GAS biological  
 activity indicates polynucleotides and polypeptides corresponding to this gene are useful  
 30 for the treatment or diagnosis of hepatic conditions such as hepatocellular carcinoma.  
 Moreover, the expression within embryonic tissue and other cellular sources marked by  
 proliferating cells indicates that this protein may play a role in the regulation of cellular  
 division, and may show utility in the diagnosis and treatment of cancer and other  
 proliferative disorders. Similarly, developmental tissues rely on decisions involving cell  
 35 differentiation and/or apoptosis in pattern formation. Thus this protein may also be  
 involved in apoptosis or tissue differentiation and could again be useful in cancer

therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:68 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1571 of SEQ ID NO:68, b is an integer of 15 to 1585, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:68, and where b is greater than or equal to a + 14.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 59

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: GTRGLQNHRT (SEQ ID NO:603). Polynucleotides encoding these polypeptides are also encompassed by the invention.

20 This gene is expressed primarily in prostate, and to a lesser extent, in tonsil and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, immune, developmental, or pulmonary disorders and/or diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, immune, developmental, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, pulmonary surfactant or sputum, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:183 as residues: Lys-32 to Lys-38.

The tissue distribution in prostate indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cancers, particularly of the prostate. The expression within tonsils indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. The expression also indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. Moreover, the expression within fetal tissue indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis and treatment of cancer and other proliferative disorders. Similarly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:69 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1662 of SEQ ID NO:69, b is an integer of 15 to 1676, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:69, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 60

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: ELSGLG (SEQ ID NO:604). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in the brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, central nervous system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:184 as residues: Tyr-15 to Lys-21, Pro-62 to Phe-68.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of CNS disorders (such as Parkinson's disease). Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:70 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1330 of SEQ ID NO:70, b is an integer of 15 to 1344, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:70, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 61**

5           The gene encoding the disclosed cDNA is believed to reside on chromosome 3. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 3.

          This gene is expressed primarily in the brain.

          Therefore, polynucleotides and polypeptides of the invention are useful as  
10   reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, CNS diseases, such as Alzheimers and Parkinson's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a  
15   number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to  
20   the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

          Preferred epitopes include those comprising a sequence shown in SEQ ID NO:185 as residues: Asp-44 to Cys-53, Asp-56 to Lys-66, Ser-78 to Lys-84.

          The tissue distribution in brain tissue indicates polynucleotides and polypeptides  
25   corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of CNS diseases such as Alzheimers and Parkinson's disease. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Huntington's Disease, Tourette Syndrome,  
30   meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance,  
35   and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition,

homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:71 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1460 of SEQ ID NO:71, b is an integer of 15 to 1474, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:71, and where b is greater than or equal to a + 14.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 62

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: MDDIKI (SEQ ID NO:605), NFCVSKNTFN RVKRPIKWVKIF ANDISCKRLISRIHKEILPFNNKKQPDFKVKKS RK (SEQ ID NO:606), FNRV KRPIKWVKIFANDISCKRLISRIHKE (SEQ ID NO:607), ETQMANKYMKRCSTL (SEQ ID NO:608), VIRELQVKATRRCHYTPIKWSKSKFLISSNADEYVEPTRTLI HCWWKCKIVQPLCKTAW (SEQ ID NO:609), and/or ATRRCHYTPIKWSKSKT LISSN (SEQ ID NO:610). Polynucleotides encoding these polypeptides are also encompassed by the invention.

25

This gene is expressed primarily in duodenum, and to a lesser extent, in brain frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, gastrointestinal, neural, or endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., gastrointestinal, neural, endocrine, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal

35



fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in colon indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of some gastrointestinal disorders, particularly cancers. Moreover, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states, behavioral disorders, or inflammatory conditions which include, but are not limited to Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates that it plays a role in normal neural function. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:72 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1998 of SEQ ID NO:72, b is an integer of 15 to 2012, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:72, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 63

This gene is expressed primarily in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, leukemia, immune deficiency syndromes, and other immune related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases of bone marrow, such as leukemia, bone cancer and immune deficiency syndrome. Furthermore, the polypeptides or polynucleotides are also useful to enhance or protect proliferation, differentiation, and functional activation of hematopoietic progenitor cells (e.g., bone marrow cells), useful in treating cancer patients undergoing chemotherapy or patients undergoing bone marrow transplantation. The uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:73 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1253 of SEQ ID NO:73, b is an integer of 15 to 1267, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:73, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

When tested against sensory neuron cell lines, supernatants removed from cells containing this gene activated the EGR1 assay. Thus, it is likely that this gene activates sensory neuronal cells through a signal transduction pathway. Early growth response 1 (EGR1) is a promoter associated with certain genes that induces various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation.

This gene is expressed primarily in the testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive system-related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., testes, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in testes indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of reproductive system-related diseases. Furthermore, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that may be expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product may be expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:74 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1734 of SEQ ID NO:74, b is an integer of 15 to 1748, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:74, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 65

This gene is expressed primarily in synovial fibroblasts.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, rheumatoid arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The restricted tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of rheumatoid arthritis, since synovial fibroblasts are associated with the synovium and cartilage. In addition, the expression of this gene product in synovium indicates a role in the detection and treatment of disorders and conditions afflicting the skeletal system, in particular osteoporosis as well as disorders afflicting connective tissues (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation), such as in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and

specific joint abnormalities as well as chondrodysplasias (ie. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:75 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1556 of SEQ ID NO:75, b is an integer of 15 to 1570, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:75, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 66

This gene is expressed primarily in ovarian cancer, and to a lesser extent in fetal tissues such as fetal liver and fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, particularly of the ovary. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., ovary, fetal, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:190 as residues: Pro-28 to Gln-33.

The tissue distribution in ovarian cancer tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of cancers; e.g., ovarian cancer, as well as other tissues where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:76 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 510 of SEQ ID NO:76, b is an integer of 15 to 524, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:76, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 67

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, male reproductive or hormonal related disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., testes, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:191 as residues: Pro-68 to Asp-73, Gln-92 to Glu-107, Gln-120 to Lys-126.

The tissue distribution in testes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of male

reproductive or hormonal disorders. Furthermore, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that may be expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product may be expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:77 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1292 of SEQ ID NO:77, b is an integer of 15 to 1306, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:77, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 68

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in human tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., tonsils, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders. Expression of this gene product in tonsils indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:78 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more



polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1465 of SEQ ID NO:78, b is an integer of 15 to 1479, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:78, and where b is greater than or equal to a + 14.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 69

This gene is expressed primarily in human thymus and six week old human embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endocrine diseases and leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., endocrine, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in thymus and developing embryonic tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of leukemia or other immune diseases, especially those which are involved in fetal development. Furthermore, the tissue distribution in thymus and developing embryonic tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ

ID NO:79 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more

- 5 polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1780 of SEQ ID NO:79, b is an integer of 15 to 1794, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:79, and where b is greater than or equal to a + 14.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 70.

This gene is expressed primarily in adult pulmonary tissue.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases of the cardiopulmonary system including asthma, bronchitis, apnea, enlarged heart, arrhythmia, strokes and heart attacks. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes
- 20 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiopulmonary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., pulmonary, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell
- 25 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:194 as residues: Pro-27 to Leu-41.

- 30 The tissue distribution in pulmonary tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment or diagnosis of diseases such as arrhythmia, apnea, asthma and possibly for the early detection and prevention of patients likely to have strokes or heart attacks. Furthermore, the tissue distribution in pulmonary tissues indicates polynucleotides and polypeptides
- 35 corresponding to this gene are useful for the detection and treatment of disorders associated with developing lungs, particularly in premature infants where the lungs are the last tissues to develop. Additionally, the tissue distribution indicates polynucleotides

and polypeptides corresponding to this gene are useful for the diagnosis and intervention of lung tumors. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

- 5 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:80 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.
- 10 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1266 of SEQ ID NO:80, b is an integer of 15 to 1280, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:80, and where b is greater than or equal to a + 14.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 71

- 20 When tested against K562 leukemia cell lines, supernatants removed from cells containing this gene activated the ISRE assay. Thus, it is likely that this gene activates leukemia cells through the Jak-STAT signal transduction pathway. The interferon-sensitive response element is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells.
- 25 Therefore, activation of the Jak-STAT pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Furthermore, contact of cells with supernatant expressing the product of this gene increases the permeability of THP-1 monocyte cells to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated
- 30 when the product of this gene binds a receptor on the surface of the monocyte cell. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocyte cells.

This gene is expressed primarily in adult human spleen and adult human testis.

- 35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., spleen, testes, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:195 as residues: Pro-32 to Gly-39.

The tissue distribution in spleen, in addition to the biological activity data, indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of immune disorders. Furthermore, this gene may play a role in the survival, proliferation, and/or differentiation of hematopoietic cells in general, and may be of use in the augmentation of the numbers of stem cells and committed progenitors. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:81 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 960 of SEQ ID NO:81, b is an integer of 15 to 974, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:81, and where b is greater than or equal to a + 14.

30

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 72**

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

35 ELGLVIITAWIILCHSSSKNPVGGRIQLAIAIVITLFPFISWVYIYINKEMRSSWP  
THCKTVI (SEQ ID NO:611), QCPQGTETEAGVSVPPRKEGGGPYVAGLTAPHVA  
GLTAPRRVLRAMAPALWRACNGL (SEQ ID NO:612), HSSSKNPVGGRIQLA

IAIVITLFPFISWVYIY (SEQ ID NO:613), and/or RKEGGGPYVAGLTAPHVA  
GLTAPRRVLRAMAP (SEQ ID NO:614). Polynucleotides encoding these  
polypeptides are also encompassed by the invention.

5 This gene is expressed primarily in liver tissues, and to a lesser extent in t-cell  
lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, hepatitis, sclerosis of the liver and cancer of the liver. Similarly,  
10 polypeptides and antibodies directed to these polypeptides are useful in providing  
immunological probes for differential identification of the tissue(s) or cell type(s). For a  
number of disorders of the above tissues or cells, particularly of the immune system,  
expression of this gene at significantly higher or lower levels may be routinely detected  
in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily  
15 fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or  
cell sample taken from an individual having such a disorder, relative to the standard  
gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
individual not having the disorder.

The tissue distribution in liver indicates polynucleotides and polypeptides  
20 corresponding to this gene are useful for the diagnosis and possible treatment of  
diseases of the liver. Since it is primarily found in the liver, and with the additional  
expression seen in T-cells, it most likely deals with the immune response in the liver,  
for example to diseases like hepatitis, sclerosis and hepatocellular carcinoma. More  
generally, the tissue distribution in liver indicates polynucleotides and polypeptides  
25 corresponding to this gene are useful for the detection and treatment of liver disorders  
and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and  
conditions that are attributable to the differentiation of hepatocyte progenitor cells).  
Protein, as well as, antibodies directed against the protein may show utility as a tumor  
marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available  
and accessible through sequence databases. Some of these sequences are related to SEQ  
ID NO:82 and may have been publicly available prior to conception of the present  
invention. Preferably, such related polynucleotides are specifically excluded from the  
scope of the present invention. To list every related sequence is cumbersome.  
35 Accordingly, preferably excluded from the present invention are one or more  
polynucleotides comprising a nucleotide sequence described by the general formula of  
a-b, where a is any integer between 1 to 1941 of SEQ ID NO:82, b is an integer of 15

to 1955, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:82, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 73

The gene encoding the disclosed cDNA is thought to reside on chromosome 10. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 10.

10 This gene is expressed primarily in myeloid progenitor cells, and to a lesser extent in leukemic cells and eosinophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, leukemia and other blood diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution in immune tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of leukemia. Furthermore, the polypeptides or polynucleotides are also useful to enhance or protect proliferation, differentiation, and functional activation of hematopoietic progenitor cells (e.g., bone marrow cells), useful in treating cancer patients undergoing chemotherapy or patients undergoing bone marrow transplantation. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

35 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:83 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 624 of SEQ ID NO:83, b is an integer of 15 to 638, where both a and b correspond to the positions of nucleotide residues shown in  
5 SEQ ID NO:83, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 74

10 The translation product of this gene shares sequence homology with "neurogenic secreted signaling protein (brn)" (see gil1150971) from Drosophila melanogaster which is thought to be important in the normal development of brain tissue. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: PGRPTRPAXAGLSSGGAAQEAPQADPRPWLAR (SEQ ID  
15 NO:615). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in the placenta and early embryonic tissue. Northern data has demonstrated that this gene is expressed in brain, stomach and colorectal adenocarcinoma.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, several types of disorders of the brain including epilepsy, mood disorders, any of a variety of types of mental retardation, and addictive disorders  
25 including alcoholism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., brain,  
30 stomach, colon, placental, embryonic, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:198 as residues: Gln-37 to Ala-42, Thr-51 to Ala-57, Pro-71 to His-79, Glu-124 to Arg-137, Ser-151 to Val-159.

The tissue distribution and homology to *Drosophila melanogaster* putative neurogenic secreted signaling protein (brn) indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment of various brain disorders as well as pre-natal testing for neuropathological conditions, such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders. Furthermore, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders of the placenta. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product may be produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product may be produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:84 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 845 of SEQ ID NO:84, b is an integer of 15 to 859, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:84, and where b is greater than or equal to a + 14.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 75**

5 The translation product of this gene shares sequence homology with a fat-specific secreted protein.

This gene is expressed primarily in the epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic disorders and male infertility. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., epididymus, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:199 as residues: Tyr-21 to Asp-40, Ser-58 to Arg-64, Thr-71 to Ser-76, Ser-106 to Thr-112.

Homology to a fat-specific gene indicates that this gene may also play a role in the treatment and/or detection of metabolic disorders such as obesity, diabetes, anorexia nervosa and bulimia. In addition, its expression primarily in the epididymus indicates a role in the treatment/detection of male fertility disorders such as infertility, low sperm count, spermatorrhea and spermiation. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:85 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

35 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1115 of SEQ ID NO:85, b is an integer of 15

to 1129, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:85, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 76

The translation product of this gene shares sequence homology with Slit, a secreted Drosophila protein which plays a role in the development of axon pathway development in the central nervous system. The Slit protein is necessary for the normal  
10 development of the midline of the CNS, particularly the midline glial cells, and for the concomitant formation of the commissural axon pathways. The process is dependent on the level of SLIT protein expression. It appears that the SLIT protein is excreted by the midline glial cells, where it is synthesised and is eventually associated with the surfaces of axons that traverse them. Contact of cells with supernatant expressing the  
15 product of this gene increases the permeability of THP-1 monocyte cells to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product of this gene binds a receptor on the surface of the monocyte cell. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocyte cells. Furthermore, when tested against  
20 U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in  
25 the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in infant brain, and to a lesser extent in adult cerebellum and frontal cortex.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
35 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell

types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:200 as residues: Glu-25 to Lys-33, Glu-115 to Lys-120.

The tissue distribution primarily in brain and homology to Slit, a gene involved in axon pathway development, indicates polynucleotides and polypeptides

corresponding to this gene are useful for the treatment/detection of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, spinal cord injury, brain injuries, crushed (optic) nerve, amyotrophic lateral sclerosis, diabetes caused nerve damage, strokes, epilepsy, multiple sclerosis, paraplegia retinal degeneration, Huntingtons Disease, facial nerve damage, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:86 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2660 of SEQ ID NO:86, b is an integer of 15 to 2674, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:86, and where b is greater than or equal to a + 14.

### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 77

The translation product of this gene shares sequence homology with human endothelial cell multimerin, which is a secreted protein that binds to the extracellular matrix and is thought to be involved in hemostasis. Multimerin is a factor V/Va-binding protein and may function as a carrier protein for platelet factor V (J. Biol Chem 1995 Aug 4;270(31):18246-51 ). Contact of cells with supernatant expressing the product of this gene increases the permeability of THP-1 Monocyte cells to calcium. Thus, it is

likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product of this gene binds a receptor on the surface of the monocyte cell. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocyte cells.

- 5           This gene is expressed primarily in a variety of hematopoietic cells including T-cells, dendritic cells and B-cells as well as cells and tissues of epithelial and endothelial origin including healing wounds and keratinocytes, as well as placenta.

- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, acute internal injury, blood clotting disorders and other disorders of hemostasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hemostasis, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., endothelial, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 10  
15  
20

          Preferred epitopes include those comprising a sequence shown in SEQ ID NO:201 as residues: Ala-43 to Trp-57, Ser-81 to Gly-88, Tyr-125 to Asp-134, Pro-141 to Gly-154, Val-172 to Glu-178, Lys-296 to Gly-305, Leu-307 to Arg-314, Thr-335 to His-341.

- 25           The tissue distribution in endothelial tissues, and the homology to human endothelial cell multimerin, indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders involving the vasculature. Elevated expression of this gene product by endothelial cells indicates that it may play vital roles in the regulation of endothelial cell function; secretion; proliferation; or angiogenesis. Furthermore, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders of the placenta. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product may be produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene
- 30  
35

product may be produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells, as supported by the biological activity data mentioned previously. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:87 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1622 of SEQ ID NO:87, b is an integer of 15 to 1636, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:87, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 78

In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

HYXSTPGRVPVRQFAAASTSGGPWVPGGXLEAPFQVAPSLSHSTPVFPGLI (SEQ ID NO:616). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in osteoblasts.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, degenerative conditions of the bone including arthritis and osteoporosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., skeletal, cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:202 as residues: Thr-45 to Cys-50, Met-55 to Pro-60.

The tissue distribution in osteoblasts indicates polynucleotides and polypeptides corresponding to this gene are useful for treating degenerative conditions of the bone mediated by alterations in the activity ratio of osteoblasts and osteoclasts. Furthermore, elevated levels of expression of this gene product in osteoblastoma indicates that it may play a role in the survival, proliferation, and/or growth of osteoblasts. Therefore, it may be useful in influencing bone mass in such conditions as osteoporosis. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

- 15 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:88 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.
- 20 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1625 of SEQ ID NO:88, b is an integer of 15 to 1639, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:88, and where b is greater than or equal to a + 14.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 79

- The gene encoding the disclosed cDNA is thought to reside on chromosome 17.
- 30 Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 17. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: ARGKYESAQPGGTQPEPGLGAR (SEQ ID NO:617). Polynucleotides encoding these polypeptides are also encompassed by the invention.

- 35 This gene is expressed primarily in pituitary, cerebellum and kidney and to a lesser extent in a range of fetal tissues including lung, heart and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic, neurological, and renal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine, renal and nervous systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., kidney, fetal, brain, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:203 as residues: Pro-29 to Gly-34, Gln-79 to Arg-84, Arg-146 to Arg-152, Ser-183 to Ser-193, Gly-233 to His-241, Tyr-265 to Pro-278, Thr-304 to Arg-320, Leu-328 to Gly-333, Glu-385 to Arg-399.

The high expression of a secreted gene in the pituitary indicates a role for this gene or gene product in the treatment/detection of metabolic disorders associated with the endocrine system, such as growth and developmental defects. Expression in the cerebellum indicates a role in the treatment/detection of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. Expression in the kidney indicates a role in the treatment/detection of renal disorders such as kidney failure, Wilms Tumor and kidney stones, as well as nephritis, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:89 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more

polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1846 of SEQ ID NO:89, b is an integer of 15 to 1860, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:89, and where b is greater than or equal to a + 14.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 80

The translation product of this gene shares sequence homology with ras-related proteins in rats which is thought to be involved in cellular signaling.

This gene is expressed primarily in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:204 as residues: Met-40 to Thr-46, Ala-57 to Glu-64, Ser-85 to Leu-91.

The tissue distribution in immune system tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility



in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Expression of this gene product in T cells also strongly indicates a role for this protein in immune function and immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:90 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 825 of SEQ ID NO:90, b is an integer of 15 to 839, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:90, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 81

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in fetal liver, osteoclastoma and neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases of the bone, haemopoietic system and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune,

haemopoietic and bone, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, skeletal, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal liver, osteoclastoma and neutrophils indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases of the bone, haemopoietic and immune systems, as well as cancer. Furthermore, elevated levels of expression of this gene product in osteoclastoma indicates that it may play a role in the survival, proliferation, and/or growth of osteoclasts. Therefore, it may be useful in influencing bone mass in such conditions as osteoporosis. More generally, as evidenced by expression in fetal liver/spleen, as well as the biological activity data, this gene may play a role in the survival, proliferation, and/or differentiation of hematopoietic cells in general, and may be of use in the augmentation of the numbers of stem cells and committed progenitors. Expression of this gene product in neutrophils also indicates that it may play a role in mediating responses to infection and controlling immunological responses, such as those that occur during immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:91 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1131 of SEQ ID NO:91, b is an integer of 15 to 1145, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:91, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 82

This gene is expressed primarily in endometrial stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, female infertility. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., reproductive, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:206 as residues: Gln-26 to Asn-51.

The tissue distribution in endometrial cells indicates polynucleotides and polypeptides corresponding to this gene are useful for treating female infertility. The protein product is likely involved in preparation of the endometrium of implantation and could be administered either topically or orally. Alternatively, this gene could be transfected in gene-replacement treatments into the cells of the endometrium and the protein products could be produced. Similarly, these treatments could be performed during artificial insemination for the purpose of increasing the likelihood of implantation and development of a healthy embryo. In both cases this gene or its gene product could be administered at later stages of pregnancy to promote healthy development of the endometrium. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:92 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2036 of SEQ ID NO:92, b is an integer of 15 to 2050, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:92, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 83**

5           The gene encoding the disclosed cDNA is thought to reside on chromosome 19. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 19.

          This gene is expressed primarily in tissues of the central nervous system (predominantly the cerebellum) and immune system (predominantly the tonsils).

10           Nucleic acids of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the immune system and CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the  
15 tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and neurological system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, neurological, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell  
20 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

          Preferred epitopes include those comprising a sequence shown in SEQ ID NO:207 as residues: Pro-43 to Leu-49, Pro-61 to Gly-66, Ser-71 to Ser-83.

25           The tissue distribution in the immune system indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Expression of this gene product in tonsils indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This  
30 gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed  
35 tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In

addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Furthermore, the tissue distribution in the central nervous system indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:93 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1159 of SEQ ID NO:93, b is an integer of 15 to 1173, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:93, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 84

This gene is expressed primarily in testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive disorders and testes diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the testes, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., testes, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:208 as residues: Lys-28 to His-35, Asn-58 to Gly-64, Thr-80 to Asn-86, Pro-96 to Glu-111, Pro-124 to Phe-133.

The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of testes disorders. Furthermore, the tissue distribution in testes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that may be expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product may be expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:94 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 808 of SEQ ID NO:94, b is an integer of 15 to 822, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:94, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 85**

5 This gene is expressed primarily in infant brain, bone marrow and activated T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental, immune and hematopoietic disorders. Similarly, 10 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, hematopoietic and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, 15 developmental, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:209 as residues: Asn-23 to Val-37.

Expression in infant brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of mental retardation and other developmental disorders in addition to neurodegenerative disease states and 25 behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. Expression of the gene in bone marrow and in B-cells indicates a role in the treatment and/or detection of immune disorders such as arthritis, asthma, immunodeficiency diseases and leukemia. Protein, as well as, antibodies directed 30 against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:95 and may have been publicly available prior to conception of the present 35 invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more

polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1063 of SEQ ID NO:95, b is an integer of 15 to 1077, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:95, and where b is greater than or equal to a + 14.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 86

The gene encoding the disclosed cDNA is thought to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence: SCGSSRRSAKRSLTLKLIDFSHRI (SEQ ID NO:618). Polynucleotides encoding these polypeptides are also encompassed by the invention.

15 This gene is expressed primarily in infant brain, fetal liver and fetal spleen, and to a lesser extent in macrophages, T-cells, erythroid cells and myeloid progenitor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological, developmental and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, neurological, developing, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:210 as residues: Val-34 to Leu-48, Val-51 to Gly-67, Lys-74 to Asp-81, Thr-93 to Glu-98, Ser-138 to His-149, Ala-186 to Gln-201, Pro-257 to Arg-271.

Expression in infant brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of mental retardation and other developmental disorders in addition to neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons

35



Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. Its distribution in fetal liver and fetal spleen indicates that this gene may play a role in the development of the hematopoietic and immune systems and that it may play a role in the treatment/detection of immune system disorders such as leukemia, arthritis and asthma. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:96 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2078 of SEQ ID NO:96, b is an integer of 15 to 2092, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:96, and where b is greater than or equal to a + 14.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 87

This gene is expressed primarily in hippocampus, and to a lesser extent in fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, any of a variety of brain disorders including epilepsy, stroke, palsy, and mood disorders including unipolar and bipolar depression. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., brain, heart, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in hippocampus indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, epilepsy, stroke, palsy, and mood disorders including unipolar and bipolar depression, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:97 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1338 of SEQ ID NO:97, b is an integer of 15 to 1352, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:97, and where b is greater than or equal to a + 14.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 88

The translation product of this gene shares sequence homology with a C. elegans protein (coded for by C. elegans cDNA yk112f3.5). The gene encoding the disclosed cDNA is thought to reside on chromosome 3. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 3. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

HYFLRTVSGLSVVPVSLRCCMCPPTGAPATAHSPFDPPALPIQFEYQQA  
(SEQ ID NO:619), QLEAEIENLSWKVERADSYDRGDLENQMHIAEQRRRT  
35 LLKDFHDT (SEQ ID NO:620), VPVSLRCCMCPPTGAPATAHS (SEQ ID  
NO:621), and/or SWKVERADSYDRGDLENQMHIAEQR (SEQ ID NO:622).

Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, congenital disorders of the liver and spleen. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, hepatic, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal liver and spleen indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g. hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). More generally, as evidenced by expression in fetal liver/spleen, this gene may play a role in the survival, proliferation, and/or differentiation of hematopoietic cells in general, and may be of use in augmentation of the numbers of stem cells and committed progenitors. Expression of this gene product in primary dendritic cells also indicates that it may play a role in mediating responses to infection and controlling immunological responses, such as those that occur during immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:98 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of  $a-b$ , where  $a$  is any integer between 1 to 899 of SEQ ID NO:98,  $b$  is an integer of 15 to 913, where both  $a$  and  $b$  correspond to the positions of nucleotide residues shown in SEQ ID NO:98, and where  $b$  is greater than or equal to  $a + 14$ .

**FEATURES OF PROTEIN ENCODED BY GENE NO: 89**

5           This gene is expressed primarily in neutrophils.

          Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, acute immunological disorders such as inflammation. Similarly,  
10   polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily  
15   fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

          The tissue distribution in neutrophils indicates polynucleotides and polypeptides  
20   corresponding to this gene are useful for treating an acute inflammatory response. Furthermore, this gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies  
25   directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion  
30   of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Expression of this gene product in neutrophils also strongly indicates a role for this protein in immune function and immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed  
35   tissues.

          Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ

ID NO:99 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more  
5 polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 707 of SEQ ID NO:99, b is an integer of 15 to 721, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:99, and where b is greater than or equal to a + 14.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 90

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates  
15 myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS  
20 element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in fetal liver and fetal spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
25 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental and/or immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and developmental systems,  
30 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., liver, spleen, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
35 fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:214 as residues: His-23 to Leu-31, His-33 to Pro-41.

The distribution of this gene in fetal liver and fetal spleen and the biological activity data indicates it may play a role in the development of the immune and hematopoietic systems. It may, therefore, play a role in the treatment and/or detection of immune and/or hematopoietic disorders including leukemia, arthritis and asthma.

- 5 Furthermore, this gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for
- 10 immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the
- 15 differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

- Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ
- 20 ID NO:100 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of
- 25 a-b, where a is any integer between 1 to 631 of SEQ ID NO:100, b is an integer of 15 to 645, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:100, and where b is greater than or equal to a + 14.

### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in brain and osteoclastoma to a lesser extent in placenta.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological, bone and reproductive disorders. Similarly, polypeptides

and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, bone and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., skeletal, brain, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:215 as residues: Phe-47 to Cys-54.

Expression of this gene in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. Furthermore, expression in osteoclastoma indicates a role in the treatment and/or detection of bone damage such as fractures and dislocations. Elevated levels of expression of this gene product in osteoclastoma indicates that it may play a role in the survival, proliferation, and/or growth of osteoclasts. Therefore, it may be useful in influencing bone mass in such conditions as osteoporosis. Expression in the placenta indicates a role in the treatment and/or detection of pregnancy disorders such as miscarriage, birth defects, premature birth, in addition to disorders such as placenta previa and placentitis. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product may be produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product may be produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:101 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 549 of SEQ ID NO:101, b is an integer of 15 to 563, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:101, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 92

This gene is expressed primarily in T-cells, bone marrow, fetal liver/spleen and and to a lesser extent in adipocytes, kidney, melanocytes and stimulated fibroblasts.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disease characterized by alterations in T cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in immune tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for treating autoimmune diseases or proliferative disorders of the developing immune system. This gene product is primarily expressed in hematopoietic cells and tissues, suggesting that it plays a role in the survival, proliferation, and/or differentiation of hematopoietic lineages. This is particularly supported by the expression of this gene product in fetal liver and bone marrow, the two primary sites of definitive hematopoiesis. Expression of this gene product in T cells also strongly indicates a role for this protein in immune function and



immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:102 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1310 of SEQ ID NO:102, b is an integer of 15 to 1324, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:102, and where b is greater than or equal to a + 14.

15

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 93**

The gene encoding the disclosed cDNA is thought to reside on chromosome 4. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 4.

This gene is expressed primarily in fetal tissues including fetal liver/spleen and to a lesser extent in lung, bone marrow, adrenal gland tumor and in the Ntera2 cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adrenal gland or lungs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., developing tissues, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35

The tissue distribution in fetal and developing tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for treating tumors formed by

poorly differentiated cells, as well as tumors of other tissues where expression has been observed. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available  
5 and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:103 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more  
10 polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1717 of SEQ ID NO:103, b is an integer of 15 to 1731, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:103, and where b is greater than or equal to a + 14.

15

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 94**

This gene is expressed primarily in the prostate derived cell line PC3 and fetal liver/spleen.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostatic hypertrophy or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes  
25 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., prostate, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell  
30 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:218 as residues: Leu-26 to Ser-33.

35 The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the prostate including prostatic tumors or benign prostatic hypertrophy. Protein, as well as, antibodies

directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:104 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1452 of SEQ ID NO:104, b is an integer of 15 to 1466, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:104, and where b is greater than or equal to a + 14.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in small intestine, and to a lesser extent in breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases of the small intestine. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the small intestine, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., small intestine, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:219 as residues: Glu-37 to Gly-45.

The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of diseases involving the small intestine, such as cancer of the small intestine or other tissues where expression has been indicated. Protein, as well as, antibodies directed against the

protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:105 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1289 of SEQ ID NO:105, b is an integer of 15 to 1303, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:105, and where b is greater than or equal to a + 14.

## 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

This gene is expressed primarily in fast-growing tissues such as tumor and fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, growth disorders such as tumorigenesis and growth retardation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fast-growing tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., rapidly proliferating cells, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:220 as residues: Phe-32 to Cys-37.

The tissue distribution in rapidly-proliferating tissues and cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of growth disorders. Furthermore, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are

useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:106 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1502 of SEQ ID NO:106, b is an integer of 15 to 1516, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:106, and where b is greater than or equal to a + 14.

20

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 97**

When tested against Jurkat T-cells and U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates both T-cells and myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in placenta, and to a lesser extent in the endometrium.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, pregnancy disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., placental, reproductive, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5      Expression of this gene in the placenta and endometrium indicates a role in the treatment and/or detection of pregnancy disorders such as miscarriage, birth defects, premature birth, in addition to disorders such as endometriosis, placenta previa and placentitis. Furthermore, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders of the placenta. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product may be produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product may be produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. Alternatively, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for treating female infertility. The protein product is likely involved in preparation of the endometrium of implantation and could be administered either topically or orally. Alternatively, this gene could be transfected in gene-replacement treatments into the cells of the endometrium and the protein products could be produced. Similarly, these treatments could be performed during artificial insemination for the purpose of increasing the likelihood of implantation and development of a healthy embryo. In both cases this gene or its gene product could be administered at later stages of pregnancy to promote healthy development of the endometrium. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:107 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1675 of SEQ ID NO:107, b is an integer of 15 to 1689, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:107, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 98

The gene encoding the disclosed cDNA is thought to reside on chromosome 5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5. Recently another group gened and sequenced this gene, calling it MDC-3.13 isoform 1 (Genbank Accession Number: g3860095), which is believed to be a cellular factor involved in the differentiation of dendritic cells. In specific embodiments, polypeptides of the invention comprise the following amino acid sequence:

HEAWLRSAGTREPPREQRTRRRQTAQLALQVPAPSRTPPMATDVFNSKNLAVX  
 AQKKILGKMVSKSIATTLIDDSSEVLDELYRVTREYTQNKKEAEKIIKNLIKTVI  
 KLAILYRNNQFNQDELALMEKFKKKVHQLAMTVVSFHQVDYTFDRNVLSRL  
 25 NECREMLHQIIQRHLTAKSHGRVNNVFDHFSDCEFLAALYNPFGNFKPHLQKL  
 CDGINKMLDEENI (SEQ ID NO:623), HEAWLRSAGTREPPREQRTRRRQTAQLA  
 LQVPAPSRTPPMATDVFNSKNLAV (SEQ ID NO:624), XAQKKILGKMVSKSIAT  
 TLIDDSSEVLDELYRVTREYTQNKKEAEKII (SEQ ID NO:625), KNLIKTVIKLA  
 ILYRNNQFNQDELALMEKFKKKVHQLAMTVVSFHQVDYTF (SEQ ID NO:626),  
 30 DRNVLSRLNECREMLHQIIQRHLTAKSHGRVNNVFDHFSDCEFLAALYNPF  
 (SEQ ID NO:627), and/or GNFKPHLQKLCDGINKMLDEENI (SEQ ID NO:628).  
 Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in placenta, spleen from CLL patients and various T cell libraries, and to a lesser extent in lung, bone marrow, neutrophil, osteoclastoma, and lymphoma tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases of the blood particularly diseases afflicting T cells and tumors of blood cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., placental, immune, vascular, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in immune tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the blood including leukemias, lymphomas and diseases that alter T-cell function or proliferation. Furthermore, the tissue distribution in placenta indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders of the placenta. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product may be produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product may be produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:108 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more



polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1929 of SEQ ID NO:108, b is an integer of 15 to 1943, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:108, and where b is greater than or equal to a + 14.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in rejected kidney, placenta, and melanocytes.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, acute or chronic renal failure. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for  
15 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., renal, placental, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample  
20 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:223 as residues: Thr-41 to Pro-47.

25 The tissue distribution in kidney indicates polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the kidney, including renal failure of either an acute or chronic nature, as well as nephritis, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to  
30 Wilms Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

35 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:109 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the

scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1580 of SEQ ID NO:109, b is an integer of 15  
5 to 1594, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:109, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 100

10

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which  
15 are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

20

This gene is expressed primarily in spinal cord, and to a lesser extent in melanocytes and fetal spleen/liver.

25

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
25 not limited to, central nervous system diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues  
30 or cell types (e.g., central nervous system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35

The tissue distribution in spinal cord tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of central nervous system disorders, such as Alzheimers Disease, Parkinsons Disease,

Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:110 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1728 of SEQ ID NO:110, b is an integer of 15 to 1742, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:110, and where b is greater than or equal to a + 14.

20

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 101**

This gene is expressed primarily in breast and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, breast related disorders and inflammatory diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast tissue and dendritic cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., breast, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in breast indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of breast related diseases and inflammatory disorders. Furthermore, the tissue distribution in breast indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and intervention of breast tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:111 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1487 of SEQ ID NO:111, b is an integer of 15 to 1501, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:111, and where b is greater than or equal to a + 14.

20

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 102

When tested against sensory neuron cell lines, supernatants removed from cells containing this gene activated the EGR1 assay. Thus, it is likely that this gene activates neuronal cells through a signal transduction pathway. Early growth response 1 (EGR1) is a promoter associated with certain genes that induces various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation. Furthermore, when tested against both Jurkat T-cells and U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates both T-cells and myeloid cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, synovial sarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the synovial sarcoma, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., synovium, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in synovial sarcoma, and the biological activity data, suggest that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of synovial sarcoma. In general, the expression of this gene product in synovium indicates a role in the detection and treatment of disorders and conditions afflicting the skeletal system, in particular osteoporosis as well as disorders afflicting connective tissues (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation), such as in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias (ie. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:112 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 777 of SEQ ID NO:112, b is an integer of 15 to 791, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:112, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 103**

This gene is expressed primarily in human tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
5 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, relating to inflammatory diseases such as tonsilitis, and immune system  
disorders. Similarly, polypeptides and antibodies directed to these polypeptides are  
10 useful in providing immunological probes for differential identification of the tissue(s)  
or cell type(s). For a number of disorders of the above tissues or cells, particularly of  
the immune system, expression of this gene at significantly higher or lower levels may  
be routinely detected in certain tissues or cell types (e.g., immune, cancerous and  
wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal  
15 fluid) or another tissue or cell sample taken from an individual having such a disorder,  
relative to the standard gene expression level, i.e., the expression level in healthy tissue  
or bodily fluid from an individual not having the disorder.

The tissue distribution in tonsils indicates polynucleotides and polypeptides  
corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of  
lymphoid tissue disorders such as tonsilitis. Furthermore, the tissue distribution  
20 indicates polynucleotides and polypeptides corresponding to this gene are useful for the  
diagnosis and treatment of a variety of immune system disorders. Expression of this  
gene product in tonsils indicates a role in the regulation of the proliferation; survival;  
differentiation; and/or activation of potentially all hematopoietic cell lineages, including  
blood stem cells. This gene product may be involved in the regulation of cytokine  
25 production, antigen presentation, or other processes that may also suggest a usefulness  
in the treatment of cancer (e.g. by boosting immune responses). Since the gene is  
expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies  
directed against the protein may show utility as a tumor marker and/or immunotherapy  
targets for the above listed tissues. Therefore it may be also used as an agent for  
30 immunological disorders including arthritis, asthma, immune deficiency diseases such  
as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and  
psoriasis. In addition, this gene product may have commercial utility in the expansion  
of stem cells and committed progenitors of various blood lineages, and in the  
differentiation and/or proliferation of various cell types. Protein, as well as, antibodies  
35 directed against the protein may show utility as a tumor marker and/or immunotherapy  
targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:113 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1623 of SEQ ID NO:113, b is an integer of 15 to 1637, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:113, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 104

- 15 This gene is expressed primarily in activated T-cells and prostate tissue.
- Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T lymphocytes related diseases and inflammation of the prostate.
- 20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune, reproductive, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.
- 25
- 30 Preferred epitopes include those comprising a sequence shown in SEQ ID NO:228 as residues: Arg-24 to Trp-36.
- The tissue distribution in immune system tissues and prostate tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune and reproductive disorders. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin,
- 35

the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Expression of this gene product in T cells also strongly indicates a role for this protein in immune function and immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:114 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1574 of SEQ ID NO:114, b is an integer of 15 to 1588, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:114, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 105

This gene is expressed primarily in human adult pulmonary tissue and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, relating to the lung, neurological and immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the respiratory, nervous, and immune systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., pulmonary, immune, nervous, cancerous and wounded tissues) or bodily fluids (e.g.,



serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5           The tissue distribution in pulmonary tissue and infant brain tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment for disorders relating to the pulmonary system, the central nervous system, and the immune system. Furthermore, the tissue distribution in pulmonary tissue and fetal tissue indicates polynucleotides and polypeptides
- 10 corresponding to this gene are useful for the detection and treatment of disorders associated with developing lungs, particularly in premature infants where the lungs are the last tissues to develop. The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and intervention of lung tumors, since the gene may be involved in the regulation of cell division.
- 15 Additionally, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS,
- 20 psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders. Also, this gene product may be involved in the regulation of cytokine production, antigen presentation, or other
- 25 processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma,
- 30 immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a
- 35 tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ

ID NO:115 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.

Accordingly, preferably excluded from the present invention are one or more

- 5 polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1912 of SEQ ID NO:115, b is an integer of 15 to 1926, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:115, and where b is greater than or equal to a + 14.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 106

- The translation product of this gene shares sequence homology with the KIAA0132 gene product, and also shares homology to *Drosophila melanogaster* ring canel protein. The gene encoding the disclosed cDNA is thought to reside on
- 15 chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in infant brain and B-cell lymphoma.

- Therefore, polynucleotides and polypeptides of the invention are useful as
- 20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, relating to the central nervous system and B cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a
- 25 number of disorders of the above tissues or cells, particularly of the central nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., central nervous system, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having
- 30 such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:230 as residues: Thr-31 to Trp-42, Gly-49 to His-54, Gly-68 to Glu-75, Ser-77 to Trp-89, Met-142 to Gly-148.

- 35 The tissue distribution in infant brain and B-cell lymphomas indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention for central nervous system and immune

disorders. Furthermore, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked disorders. Additionally, this gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses). Since the gene is expressed in cells of lymphoid origin, the gene or protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, leukemia, rheumatoid arthritis, inflammatory bowel disease, sepsis, acne, and psoriasis. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:116 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1049 of SEQ ID NO:116, b is an integer of 15 to 1063, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:116, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 107**

When tested against sensory neuron cell lines, supernatants removed from cells containing this gene activated the EGR1 assay. Thus, it is likely that this gene activates neuronal cells through a signal transduction pathway. Early growth response 1 (EGR1) is a promoter associated with certain genes that induces various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation. The gene encoding the disclosed cDNA is thought to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in human gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, relating to gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., gall bladder, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:231 as residues: Pro-45 to Pro-51.

The tissue distribution in gall bladder indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of gastrointestinal disorders. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:117 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of

a-b, where a is any integer between 1 to 1601 of SEQ ID NO:117, b is an integer of 15 to 1615, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:117, and where b is greater than or equal to a + 14.

5

## FEATURES OF PROTEIN ENCODED BY GENE NO: 108

This gene is expressed primarily in human whole brain.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of  
15 the above tissues or cells, particularly of the central nervous and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., brain, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard  
20 gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:232 as residues: Gln-58 to Asp-64, His-69 to Pro-76, Leu-101 to Glu-108.

The tissue distribution in brain tissue indicates polynucleotides and polypeptides  
25 corresponding to this gene are useful for the diagnosis and treatment of the central nervous system and endocrine system disorders. Furthermore, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic  
30 disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, or sexually-linked  
35 disorders. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:118 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1207 of SEQ ID NO:118, b is an integer of 15 to 1221, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:118, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 109

The translation product of this gene shares sequence homology with human translation initiation factor eIF3 p40 subunit.

This gene is expressed primarily in human adipose, human fetal spleen, and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, adipose, immune and nerve cell disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., adipose, immune, nervous, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID NO:233 as residues: Asn-27 to Ser-33, Gln-44 to Lys-50.

The tissue distribution fetal liver/spleen and dendritic cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune and nerve cell disorders. Furthermore, the tissue distribution in adipose tissue indicates polynucleotides and polypeptides corresponding

to this gene are useful for the treatment of obesity and other metabolic and endocrine conditions or disorders. Additionally, the protein product of this gene may show utility in ameliorating conditions which occur secondary to aberrant fatty-acid metabolism (e.g. aberrant myelin sheath development), either directly or indirectly. Also, the tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of hematopoietic disorders. This gene product is primarily expressed in hematopoietic cells and tissues, suggesting that it plays a role in the survival, proliferation, and/or differentiation of hematopoietic lineages. This is particularly supported by the expression of this gene product in fetal liver, which is a primary site of definitive hematopoiesis. Expression of this gene product in primary dendritic cells also strongly indicates a role for this protein in immune function and immune surveillance. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:119 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1135 of SEQ ID NO:119, b is an integer of 15 to 1149, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:119, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 110

The translation product of this gene shares sequence homology with Ig V-chain, which is thought to be important in immune function. When tested against Jurkat cell lines, supernatants removed from cells containing this gene activated the NF-kB transcription factor. Thus, it is likely that this gene activates Jurkat cells by activating a transcriptional factor found within these cells. Nuclear factor kB is a transcription factor activated by a wide variety of agents, leading to cell activation, differentiation, or apoptosis. Reporter constructs utilizing the NF-kB promoter element are used to screen supernatants for such activity.

This gene is expressed in human synovial sarcoma, infant brain INIB cells, macrophages (GM-CSF treated), human endometrial stromal cells-treated with estradiol, human pancreas tumor, hemangiopericytoma, human endometrial tumor, chronic lymphocytic leukemia and a human colon carcinoma (HCC) cell line.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers such as human synovial sarcoma, human pancreas tumor, hemangiopericytoma, human endometrial tumor, chronic lymphocytic leukemia and  
10 human colon carcinoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues or cell types (e.g., immune,  
15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred epitopes include those comprising a sequence shown in SEQ ID  
20 NO:234 as residues: Leu-21 to Ala-30, Ser-38 to Asp-47, Pro-87 to Asp-94, Leu-197 to Thr-204, Pro-256 to Ser-262, Thr-277 to Arg-282, Thr-293 to Trp-303.

The tissue distribution in numerous cancerous tissues, and the homology to Ig V-chain, as well as the biological activity data, indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of  
25 cancers, including human synovial sarcoma, human pancreas tumor, hemangiopericytoma, human endometrial tumor, chronic lymphocytic leukemia and human colon carcinoma, as well as other tissues where expression has been demonstrated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:120 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome.  
35 Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1501 of SEQ ID NO:120. b is an integer of 15



to 1515, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:120, and where b is greater than or equal to a + 14.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HFCCQ50	209463 11/14/97	Uni-ZAP XR	11	1271	1	1271	47	47	125	1	20	21	352
2	HTLAI54	209463 11/14/97	Uni-ZAP XR	12	1451	1	1451	125	125	126	1	22	23	157
3	HKABT24	209463 11/14/97	pCMVSPORT 2.0	13	2317	1	1809	66	66	127	1	22	23	553
4	HLWBF94	209463 11/14/97	pCMVSPORT 3.0	14	1472	1	1472	192	192	128	1	18	19	307
5	HFKEFF78	209463 11/14/97	Uni-ZAP XR	15	1016	1	961	92	92	129	1	18	19	166
6	HSYBG37	209463 11/14/97	pCMVSPORT 3.0	16	1239	1	1239	48	48	130	1	24	25	305
7	HTHCA77	209463 11/14/97	Uni-ZAP XR	17	1405	1	1405	160	160	131	1	24	25	219

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
8	HNHEZ51	209463 11/14/97	Uni-ZAP XR	18	1534	1	1534	106	106	132	1	32	33	98
9	HFIAX46	209463 11/14/97	pSport1	19	1233	1	1233	195	195	133	1	20	21	60
10	HFOXO72	209463 11/14/97	pSport1	20	1090	1	862	240	240	134	1	32	33	247
11	HODDW40	209463 11/14/97	Uni-ZAP XR	21	682	1	682	139	139	135	1	19	20	40
12	HSAWG42	209463 11/14/97	Uni-ZAP XR	22	770	1	770	31	31	136	1	27	28	74
13	HBMSK09	209463 11/14/97	Uni-ZAP XR	23	565	1	565	67	67	137	1	28	29	74
14	HDP AU16	209463 11/14/97	pCMV Sport 3.0	24	1356	1	1356	486	486	138	1	24	25	57
15	HFE BE12	209463 11/14/97	Uni-ZAP XR	25	617	3	617	99	99	139	1	22	23	173

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
16	HFLNB64	209463 11/14/97	Uni-ZAP XR	26	648	1	648	62	62	140	1	39	40	45
16	HCESD11	209877 05/18/98	pBluescript	121	1025	1	1025	188	188	235	1	18	19	28
17	HSAWZ41	209463 11/14/97	Uni-ZAP XR	27	1388	1	1388	98	98	141	1	24	25	57
18	HNFJF07	209463 11/14/97	Uni-ZAP XR	28	616	1	616	86	86	142	1	21	22	66
19	HNGIO57	209463 11/14/97	Uni-ZAP XR	29	828	1	828	87	87	143	1	18	19	52
20	HE7TM22	209463 11/14/97	Uni-ZAP XR	30	581	1	581	70	70	144	1	22	23	65
21	HFRBR70	209463 11/14/97	Uni-ZAP XR	31	789	1	789	40	40	145	1	20	21	56
22	HTHBK35	209463 11/14/97	Uni-ZAP XR	32	884	1	884	108	108	146	1	26	27	66

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
23	HWABA81	209463 11/14/97	pCMVSPORT 3.0	33	866	1	866	57	57	147	1	21	22	48
24	HKGAA73	209463 11/14/97	pSport1	34	1694	1	1694	38	38	148	1	27	28	88
25	HKIYP40	209463 11/14/97	pBluescript	35	1215	1	1215	43	43	149	1	32	33	76
26	HKMMW74	209463 11/14/97	pBluescript	36	1794	1	1794	202	202	150	1	21	22	41
27	HLFB127	209463 11/14/97	pBluescript SK-	37	1174	1	1174	135	135	151	1	19	20	45
28	HLQCW84	209463 11/14/97	Lambda ZAP II	38	1087	1	1087	31	31	152	1	18	19	41
29	HBNAV22	209463 11/14/97	Uni-ZAP XR	39	438	1	438	13	13	153	1	40	41	43
30	HTEAM34	209463 11/14/97	Uni-ZAP XR	40	734	1	734	63	63	154	1	28	29	122

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
31	HTHDK34	209463 11/14/97	Uni-ZAP XR	41	1346	1	1346	60	60	155	1	35	36	41
32	H6BSG32	209463 11/14/97	Uni-ZAP XR	42	998	53	998	209	209	156	1	24	25	55
33	HCFAD33	209463 11/14/97	pSport1	43	658	1	658	297	297	157	1	17	18	44
34	HDTEN81	209463 11/14/97	pCMV Sport 2.0	44	566	1	566	114	114	158	1	17	18	85
35	HFXDT43	209463 11/14/97	Lambda ZAP II	45	1277	1	1277	92	92	159	1	19	20	44
36	HNGHQ09	209463 11/14/97	Uni-ZAP XR	46	442	1	442	65	65	160	1	17	18	68
37	HHGDF16	209463 11/14/97	Lambda ZAP II	47	890	215	890	253	253	161	1	26	27	52
38	HJBCG12	209463 11/14/97	pBluescript SK-	48	737	40	737	382	382	162	1	24	25	56

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Secreted Portion	Last AA of ORF
39	HOGAW62	209463 11/14/97	pCMVSPORT 2.0	49	571	1	571	259	259	163	1	26	55
40	HSWBJ74	209463 11/14/97	pCMVSPORT 3.0	50	356	1	356	43	43	164	1	35	47
41	HGBHR26	209511 12/03/97	Uni-ZAP XR	51	913	1	913	174	174	165	1	22	129
42	HKDBF34	209511 12/03/97	pCMVSPORT 1	52	1356	1	1356	18	18	166	1	19	104
43	H6EAB28	209511 12/03/97	Uni-ZAP XR	53	1547	1	1547	116	116	167	1	21	76
44	HLWAO22	209511 12/03/97	pCMVSPORT 3.0	54	1338	1	1311	212	212	168	1	21	354
45	HAGFH53	209511 12/03/97	Uni-ZAP XR	55	2071	1	2071	96	96	169	1	36	89
46	HHENQ22	209511 12/03/97	pCMVSPORT 3.0	56	1899	1	1899	115	115	170	1	36	58

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
47	HKMLK53	209511 12/03/97	pBluescript	57	1543	1	1543	20	20	171	1	25	26	69
48	HSKGQ58	209511 12/03/97	pBluescript	58	1133	1	1133	41	41	172	1	37	38	78
49	HNFE93	209511 12/03/97	Uni-ZAP XR	59	1490	1	1480	33	33	173	1	33	34	173
50	HAIBZ39	209511 12/03/97	Uni-ZAP XR	60	1336	1	1336	48	48	174	1	17	18	63
51	HBXFP23	209511 12/03/97	ZAP Express	61	1705	178	1705	384	384	175	1	25	26	42
52	HEQBF32	209511 12/03/97	pCMVSPORT 3.0	62	1031	1	1031	97	97	176	1	25	26	113
53	HETHE81	209511 12/03/97	Uni-ZAP XR	63	1589	1	1589	182	182	177	1	23	24	155
54	HFPAC12	209511 12/03/97	Uni-ZAP XR	64	1088	1	1088	140	140	178	1	21	22	88



Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
55	HDPFF39	209511 12/03/97	pCMVSPORT 3.0	65	1256	1	1256	175	175	179	1	18	19	196
56	HFXHD88	209511 12/03/97	Lambda ZAP II	66	1602	1	1602	130	130	180	1	41	42	128
57	HFOXV65	209511 12/03/97	pSPORT1	67	938	1	938	204	204	181	1	26	27	154
58	HKADX21	209511 12/03/97	pCMVSPORT 2.0	68	1585	122	1585	414	414	182	1	18	19	106
59	HPZAB47	209511 12/03/97	pBluescript	69	1676	1	1676	34	34	183	1	18	19	47
60	HAGFE79	209511 12/03/97	Uni-ZAP XR	70	1344	1	1344	133	133	184	1	18	19	126
61	HCE1X60	209511 12/03/97	Uni-ZAP XR	71	1474	1	1474	38	38	185	1	25	26	86
62	HFXKD36	209511 12/03/97	Lambda ZAP II	72	2012	130	2012	251	251	186	1	35	36	57

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
63	HBMCU71	209511 12/03/97	pBluescript	73	1267	1	1267	77	77	187	1	21	22	68
64	HTEIV80	209511 12/03/97	Uni-ZAP XR	74	1748	1	1748	203	203	188	1	15	16	47
65	HFIAP16	209511 12/03/97	pSport1	75	1570	38	1570	44	44	189	1	30	31	50
66	HODAV86	209511 12/03/97	Uni-ZAP XR	76	524	1	524	153	153	190	1	28	29	55
67	HTEDF80	209511 12/03/97	Uni-ZAP XR	77	1306	1	1306	696	696	191	1	21	22	126
68	HTODI69	209511 12/03/97	Uni-ZAP XR	78	1479	1	1479	123	123	192	1	23	24	69
69	HE6GR02	209511 12/03/97	Uni-ZAP XR	79	1794	1	1794	100	100	193	1	18	19	70
70	HAPNY86	209511 12/03/97	Uni-ZAP XR	80	1280	1	1280	100	100	194	1	25	26	129

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
71	HTLDR33	209511 12/03/97	Uni-ZAP XR	81	974	1	974	368	368	195	1	19	20	54
72	HACBI61	209511 12/03/97	Uni-ZAP XR	82	1955	1	1955	174	174	196	1	16	17	79
73	HMEIK34	209511 12/03/97	Lambda ZAP II	83	638	1	638	243	243	197	1	24	25	41
74	HKAAK02	209551 12/12/97	pCMVSPORT 2.0	84	859	1	859	97	97	198	1	33	34	196
75	HEPAA46	209551 12/12/97	Uni-ZAP XR	85	1129	1	1129	18	18	199	1	20	21	123
76	HFPCX09	209551 12/12/97	Uni-ZAP XR	86	2674	59	2674	249	249	200	1	26	27	549
76	HFPCX09	209551 12/12/97	Uni-ZAP XR	122	2207	1	2207	185	185	236	1	26	27	66
77	HLWAA88	209551 12/12/97	pCMVSPORT 3.0	87	1636	1	1636	51	51	201	1	22	23	488

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
77	HLWAA88	209551 12/12/97	pCMVSPORT 3.0	123	1770	1	1770	35	35	237	1	22	23	113
78	HOHBV89	209551 12/12/97	pCMVSPORT 2.0	88	1639	1	1639	158	158	202	1	28	29	85
79	HHFBY69	209852 05/07/98	Uni-ZAP XR	89	1860	2	1860	71	71	203	1	25	26	399
79	HCEFL57	209551 12/12/97	Uni-ZAP XR	124	1034	1	1032	68	68	238	1	25	26	105
80	HMEKU83	209551 12/12/97	Lambda ZAP II	90	839	1	839	35	35	204	1	28	29	194
81	HOSBY40	209551 12/12/97	Uni-ZAP XR	91	1145	1	1145	89	89	205	1	30	31	56
82	HKFBH93	209551 12/12/97	ZAP Express	92	2050	174	2050	262	262	206	1	18	19	72
83	HMTAD67	209551 12/12/97	pCMVSPORT 3.0	93	1173	1	1173	306	306	207	1	19	20	84

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
84	HTEBP77	209551 12/12/97	Uni-ZAP XR	94	822	1	822	91	208	1	18	19	194
85	HE9CO69	209551 12/12/97	Uni-ZAP XR	95	1077	1	1077	161	209	1	26	27	41
86	HCACV51	209551 12/12/97	Uni-ZAP XR	96	2092	1	2092	173	210	1	31	32	281
87	HHPBI45	209551 12/12/97	Uni-ZAP XR	97	1352	1	1352	123	211	1	20	21	47
88	HLQDH79	209551 12/12/97	Lambda ZAP II	98	913	1	913	205	212	1	19	20	58
89	HNGFI67	209551 12/12/97	Uni-ZAP XR	99	721	1	721	344	213	1	21	22	42
90	HEIAC52	209551 12/12/97	Uni-ZAP XR	100	645	1	645	81	214	1	18	19	52
91	HFXKL58	209551 12/12/97	Lambda ZAP II	101	563	1	563	120	215	1	17	18	67

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
92	HMVAM60	209551 12/12/97	pSport1	102	1324	1	1324	96	96	216	1	26	27	56
93	HMVBR22	209551 12/12/97	pSport1	103	1731	1	1731	104	104	217	1	20	21	55
94	HPJCW04	209551 12/12/97	Uni-ZAP XR	104	1466	1	1466	44	44	218	1	19	20	57
95	HSIDJ81	209551 12/12/97	Uni-ZAP XR	105	1303	1	1303	8	8	219	1	22	23	58
96	HSLFU05	209551 12/12/97	Uni-ZAP XR	106	1516	1	1516	166	166	220	1	30	31	44
97	HEQAK71	209551 12/12/97	pCMVSPORT 3.0	107	1689	1	1689	198	198	221	1	17	18	44
98	HOSEQ49	209551 12/12/97	Uni-ZAP XR	108	1943	280	1935	544	544	222	1	32	33	51
99	HRAAM50	209551 12/12/97	pCMVSPORT 3.0	109	1594	1	1594	29	29	223	1	14	15	72

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 3' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
100	HSDFW45	209551 12/12/97	Uni-ZAP XR	110	1742	1	1742	118	118	224	1	19	20	70
101	HSLCQ82	209551 12/12/97	Uni-ZAP XR	111	1501	1	1501	233	233	225	1	22	23	57
102	HSSF108	209551 12/12/97	Uni-ZAP XR	112	791	1	791	125	125	226	1	34	35	58
103	HTOIW31	209551 12/12/97	Uni-ZAP XR	113	1637	1	1637	107	107	227	1	19	20	42
104	HTXKQ85	209551 12/12/97	Uni-ZAP XR	114	1588	1	1588	61	61	228	1	28	29	40
105	HUFBK08	209551 12/12/97	pSport1	115	1926	1	1926	114	114	229	1	17	18	41
106	HAIJAW31	209551 12/12/97	pcMVSPORT 3.0	116	1063	1	1063	41	41	230	1	22	23	164
107	HBJEE48	209551 12/12/97	Uni-ZAP XR	117	1615	1	1615	97	97	231	1	43	44	51

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	Signal Pep	AA of Signal Y	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
108	HBXGH74	209551 12/12/97	ZAP Express	118	1221	1	1221	188	188	232	1	20	21	25	129
109	HWBDM68	209551 12/12/97	pCMVSPORT 3.0	119	1149	1	1130	368	368	233	1	24	25	54	
110	HTPBW79	209511 12/03/97	Uni-ZAP XR	120	1515	118	1507	302	302	234	1	24	25	362	



Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

- 5           The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources  
10       using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

### Signal Sequences

- 15           Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1  
20       indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra*.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

- 25           In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results  
30       shown in Table 1.

- As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., +  
35       or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

### 10 **Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are:  
Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions,

- 5 interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

- As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be
- 10 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and
- 15 subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window
- 20 Size=500 or the length of the subject amino acid sequence, whichever is shorter.

- If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.
- 25 For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of
- 30 the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are
- 35 considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.



The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

#### Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-

60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred.

- 5 Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide  
15 fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

20

### **Epitopes & Antibodies**

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an  
25 epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

- 30 Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to  
35 about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### **Fusion Proteins**

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the

polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., *Nature* 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., *J. Biochem.* 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., *J. Molecular Recognition* 8:52-58 (1995); K. Johanson et al., *J. Biol. Chem.* 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., *Cell* 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

### Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods  
5 In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography,  
10 phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also  
15 be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production  
20 procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein  
25 after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

### 30 Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome  
35 identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat

polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be  
5 selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the  
10 polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome  
15 specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al.,  
20 "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides  
25 correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage  
30 analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease  
35 could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined.



First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991) ) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying

personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

5           The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set  
10   of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

          Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as  
15   tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more  
20   restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

          There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of  
25   unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

30           In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using  
35   DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

5 A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known  
10 in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine ( $^{125}\text{I}$ ,  $^{121}\text{I}$ ), carbon ( $^{14}\text{C}$ ), sulfur ( $^{35}\text{S}$ ), tritium ( $^3\text{H}$ ), indium ( $^{112}\text{In}$ ), and technetium ( $^{99\text{m}}\text{Tc}$ ), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

15 In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for  
20 NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example,  $^{131}\text{I}$ ,  
25  $^{112}\text{In}$ ,  $^{99\text{m}}\text{Tc}$ ), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human  
30 subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of  $^{99\text{m}}\text{Tc}$ . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The  
35 Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

#### **Biological Activities**

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

#### **Immune Activity**

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the

proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

5 A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

15 Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

20 A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Alternatively, a polypeptide or polynucleotide or other cells which can inhibit the hyperproliferative disorder, by increasing an immune response, or by increasing the activities of the hyperproliferative disorder or T-cells, hyperproliferative disorders can be increased by either enhancing an existing immune response. Alternatively, the method of treating hyperproliferative disorders

Examples of hyperproliferative disorders that can be treated by a polypeptide or polynucleotide of the present invention include, but are not limited to: tumors located in the: abdomen, bone, breast, digestive system, endocrine glands (adrenal, parathyroid, pituitary), head and neck, nervous (central and peripheral), soft tissue, spleen, thoracic, and urogenital system. Additionally, other hyperproliferative disorders can be treated by a polypeptide or polynucleotide of the present invention. Other hyperproliferative disorders include, but are not limited to: leukemias, lymphomas, myelodysplastic syndromes, myeloproliferative disorders, paraproteinemias, purpura, Waldenström's Macroglobulinemia, Gaucher's disease, hyperproliferative disease, besides neoplasia

### Disease

A polypeptide or polynucleotide of the present invention can be used as immunomodulatory agents. For example, by increasing the proliferation and differentiation of B and T cells. The immune response may be increased by increasing the production of antibodies, or by initiating a new immune response, or by initiating a new immune response with a polynucleotide of the present invention. Alternatively, without necessarily eliciting an immune response, the polypeptide or polynucleotide of the present invention can be used to treat or detect hyperproliferative disorders. Examples of viruses, include, but are not limited to: Herpesviridae, Adenoviridae, Rotaviridae, Bunyaviridae, Caliciviridae, Circoviridae, and Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Epstein-Barr virus, Herpes simplex virus, and Varicella-Zoster virus).

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly severe asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of immune cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate inflammation. For example, the polypeptide or polynucleotide of the present invention can inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., sepsis, shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia/reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., IL-1.)

### Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect

Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, 5 Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide 10 of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide 15 of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### **Regeneration**

20 A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal 25 disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and 30 skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or 35 polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue



regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

#### 15 **Chemotaxis**

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

#### **Binding Activity**

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

(antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

### **Other Activities**

10 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, 15 skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change 20 a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

25 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

### **Other Preferred Embodiments**

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

35 Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of

positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous  
5 nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of  
10 contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide  
15 sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide  
20 sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a  
nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ  
ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the  
First Amino Acid of the Signal Peptide and ending with the nucleotide at about the  
25 position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in  
Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising  
a nucleotide sequence which is at least 95% identical to the complete nucleotide  
sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under  
30 stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which  
35 comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1,  
which DNA molecule is contained in the material deposited with the American Type

Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.



Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide  
5 comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

10 Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

15 Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid  
20 sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human  
25 cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an  
30 individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of  
35 illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

5 Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For  
10 example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
15	Zap Express	pBK
	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSPORT 2.0	pCMVSPORT 2.0
	pCMVSPORT 3.0	pCMVSPORT 3.0
20	pCR®2.1	pCR®2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Altting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 25 17:9494 (1989)) and pBK (Altting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

35 Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., *Focus* 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR<sup>®</sup>2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., *Nuc. Acids Res.* 16:9677-9686 (1988) and Mead, D. et al., *Bio/Technology* 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with <sup>32</sup>P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25  $\mu$ l of reaction mixture with 0.5  $\mu$ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM  $MgCl_2$ , 0.01% (w/v) gelatin, 20  $\mu$ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is

used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

5

**Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

10

**Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P<sup>32</sup> using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

15

20

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

25

**Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on

30

35

either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

5 **Example 5: Bacterial Expression of a Polypeptide**

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as  
10 BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site  
15 (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses  
20 the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml).  
25 The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.<sup>600</sup>) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

30 Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from  
35 QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high

affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (*lacIq*). The origin of replication (*oriC*) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with *NdeI* and *XbaI*, *BamHI*, *XhoI*, or *Asp718*, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for *NdeI* (5' primer) and *XbaI*, *BamHI*, *XhoI*, or *Asp718* (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

#### 35 **Example 6: Purification of a Polypeptide from an Inclusion Body**



The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell  
5 culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a  
10 high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M  
15 NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

20 Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

25 To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted  
30 with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem

columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5  $\mu$ g of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

#### 15 **Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System**

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring

signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures,"

- 5 Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

- 10 The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

- 15 The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

- 20 Five  $\mu$ g of a plasmid containing the polynucleotide is co-transfected with 1.0  $\mu$ g of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One  $\mu$ g of BaculoGold™ virus DNA and 5  $\mu$ g of the plasmid are mixed in a sterile well of a  
25 microtiter plate containing 50  $\mu$ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10  $\mu$ l Lipofectin plus 90  $\mu$ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm  
30 tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

- After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life  
35 Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture

and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of <sup>35</sup>S-methionine and 5 µCi <sup>35</sup>S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

#### 20 **Example 8: Expression of a Polypeptide in Mammalian Cells**

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV1, HIV1 and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

- 5       The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a
- 10
- 15 chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

- Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse
- 20
- 25 DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

- A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)
- 30

- The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.
- 35

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five  $\mu$ g of the expression plasmid pC6 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

#### **Example 9: Protein Fusions**

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

5 For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that  
10 the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a  
15 heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGCCCAAATCTTCTGACAAAACTCACACATGCCCACCGTGCC  
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAAACC  
20 CAAGGACACCCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT  
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG  
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC  
AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG  
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC  
25 ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT  
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT  
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA  
GAGCAATGGGCAGCCGGAGAACAACACTACAAGACCACGCCTCCCGTGCTGG  
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA  
30 GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC  
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC  
GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

#### **Example 10: Production of an Antibody from a Polypeptide**

35 The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera

containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

5           In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., *Nature* 256:495 (1975); Köhler et al., *Eur. J. Immunol.* 6:511 (1976); Köhler et al., *Eur. J. Immunol.* 6:292 (1976); Hammerling et al., in: *Monoclonal Antibodies and T-Cell*  
10   *Hybridomas*, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at  
15   about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

          The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line  
20   (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (*Gastroenterology* 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

25           Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a  
30   mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific  
35   antibodies.



It will be appreciated that Fab and F(ab')<sub>2</sub> and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

#### **Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at  $2 \times 10^5$  cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in

- Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

- Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

- While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl<sub>2</sub> (anhyd); 0.00130 mg/L CuSO<sub>4</sub>·5H<sub>2</sub>O; 0.050 mg/L of Fe(NO<sub>3</sub>)<sub>3</sub>·9H<sub>2</sub>O; 0.417 mg/L of FeSO<sub>4</sub>·7H<sub>2</sub>O; 311.80 mg/L of KCl; 28.64 mg/L of MgCl<sub>2</sub>; 48.84 mg/L of MgSO<sub>4</sub>; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO<sub>3</sub>; 62.50 mg/L of NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O; 71.02 mg/L of Na<sub>2</sub>HPO<sub>4</sub>; .4320 mg/L of ZnSO<sub>4</sub>·7H<sub>2</sub>O; .002 mg/L of Arachidonic Acid; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H<sub>2</sub>O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H<sub>2</sub>O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H<sub>2</sub>O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H<sub>2</sub>O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B<sub>12</sub>; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine;

0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

#### **Example 12: Construction of GAS Reporter Construct**

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN- $\alpha$ , IFN- $\gamma$ , and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	<u>tyk2</u>	<u>JAKs</u> <u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>	<u>STATS</u>	<u>GAS(elements) or ISRE</u>
	<u>IFN family</u>						
5	IFN-a/B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	Il-10	+	?	?	-	1,3	
	<u>gp130 family</u>						
10	IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	Il-11(Pleiotrohic)	?	+	?	?	1,3	
	OnM(Pleiotrohic)	?	+	+	?	1,3	
	LIF(Pleiotrohic)	?	+	+	?	1,3	
	CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
15	G-CSF(Pleiotrohic)	?	+	?	?	1,3	
	IL-12(Pleiotrohic)	+	-	+	+	1,3	
	<u>g-C family</u>						
	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
20	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
	IL-15	?	+	?	+	5	GAS
25	<u>gp140 family</u>						
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
30	GM-CSF (myeloid)	-	-	+	-	5	GAS
	<u>Growth hormone family</u>						
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
35	EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
	<u>Receptor Tyrosine Kinases</u>						
	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
40	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:  
5':GCGCCTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCG  
10 AAATGATTTCCTCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTGTCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:  
5':CTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATG  
20 ATTTCCTCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC  
CTAACTCCGCCCCTCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC  
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGC  
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT  
TGCAAAAAGCTT:3' (SEQ ID NO:5)

25 With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase,  
30 alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter  
35 element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning  
5 site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules  
10 containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter  
15 construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

**Example 13: High-Throughput Screening Assay for T-cell Activity.**

The following protocol is used to assess T-cell activity by identifying factors,  
20 such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and  
25 Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMR1E-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately  
30 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to  
35 generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

- During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final  
5 concentration of  $10^7$  cells/ml. Then add 1ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

- 10 On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

- 15 Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

- After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12  
20 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

- The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples  
25 from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophane covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

- 30 As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

35 **Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**



The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^7$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting  $1 \times 10^8$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

#### **Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon

activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or  
5 differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

10 The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

15 Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

20 To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker)  
25 containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

30 Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

35 To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS

(Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count  
5 the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200  $\mu$ l of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50  $\mu$ l supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR  
10 can be used, such as 50 ng/ $\mu$ l of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### **Example 16: High-Throughput Screening Assay for T-cell Activity**

15 NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates  
the expression of genes involved in immune cell activation, control of apoptosis (NF-  
20  $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF-  $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I-  $\kappa$ B is phosphorylated and degraded, causing NF-  $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target  
25 genes. Target genes activated by NF-  $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating  
30 diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:

5 5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC  
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

10 PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

15 5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGACTTTCC  
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA  
TCCCGCCCCTAACTCCGCCCAGTTCGCCCATTTCTCCGCCCCATGGCTGACT  
AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC  
20 CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:  
3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2- promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII.

25 However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the  
30 NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described

in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### **Example 17: Assay for SEAP Activity**

- 5 As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

- 10 Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

- Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

- 20 Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### **Reaction Buffer Formulation:**

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25

28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is

incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca<sup>++</sup> concentration.

#### **Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity**

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating

tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately  
5 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from  
Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with  
100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr  
with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine  
(50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or  
10 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed  
with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000  
cells/well in growth medium and indirect quantitation of cell number through use of  
AlamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento,  
CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are  
15 used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture  
plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of  
Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium.  
Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20  
20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example  
11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH  
7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na<sub>3</sub>VO<sub>4</sub>, 2 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub>  
and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim  
(Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for  
25 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract  
filtered through the 0.45 mm membrane bottoms of each well using house vacuum.  
Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum  
manifold and immediately placed on ice. To obtain extracts clarified by centrifugation,  
the content of each well, after detergent solubilization for 5 minutes, is removed and  
30 centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many  
methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by  
determining its ability to phosphorylate a tyrosine residue on a specific substrate (a  
35 biotinylated peptide). Biotinylated peptides that can be used for this purpose include  
PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and



PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg<sub>2+</sub> (5mM ATP/50mM MgCl<sub>2</sub>), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

#### **Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

#### **Example 21: Method of Determining Alterations in a Gene**

##### **Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR

products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., *Nucleic Acids Research*, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., *Methods Cell Biol.* 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., *Genet. Anal. Tech. Appl.*, 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

25

**Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

#### **Example 23: Formulating a Polypeptide**

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2- hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

#### Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

**Example 25: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

**Example 26: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

5       The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the  
10       presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

15       The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

20       Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media  
25       from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

30       The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and,  
35       therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other



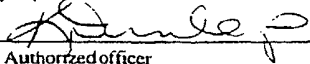
disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference. Further, the hard copy of the sequence listing submitted herewith and the corresponding computer readable form are both incorporated herein by reference in their entireties.

5

Applicant's or agent's file reference number	PZ021PCT	International application No.	unassigned
---	----------	-------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>172</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit 14 November 1997	Accession Number 209463
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application  Authorized officer	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on:  Authorized officer

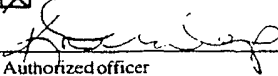
Form PCT/RO/134 (July 1992)

Applicant's or agent's file reference number	PZ021PCT	International application No	unassigned
---	----------	------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

<b>A.</b> The indications made below relate to the microorganism referred to in the description on page <u>174</u> , line <u>N/A</u>	
<b>B. IDENTIFICATION OF DEPOSIT</b> Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit 18 May 1998	Accession Number 209877
<b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
<b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if the indications are not for all designated States)	
<b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

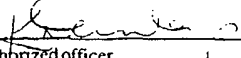
<b>For receiving Office use only</b> <input checked="" type="checkbox"/> This sheet was received with the international application  Authorized officer	<b>For International Bureau use only</b> <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer
---	--

Applicant's or agent's file reference number	PZ021PCT	International application No	unassigned
--	----------	------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>177</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit 3 December 1997	Accession Number 209511
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

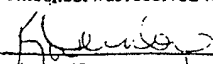
<b>For receiving Office use only</b> <input checked="" type="checkbox"/> This sheet was received with the international application  Authorized officer	<b>For International Bureau use only</b> <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer
---	--

Applicant's or agent's file reference number	PZ021PCT	International application No.	unassigned
--	----------	-------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

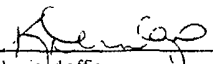
A. The indications made below relate to the microorganism referred to in the description on page <u>181</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT <span style="float: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></span>	
Name of depositary institution <u>American Type Culture Collection</u>	
Address of depositary institution (including postal code and country) <u>10801 University Boulevard</u> <u>Manassas, Virginia 20110-2209</u> <u>United States of America</u>	
Date of deposit <u>12 December 1997</u>	Accession Number <u>209551</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
<u></u> Authorized officer	_____ Authorized officer

Applicant's or agent's file reference number	PZ021PCT	International application No	unassigned
---	----------	------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>182</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution <u>American Type Culture Collection</u>	
Address of depositary institution (including postal code and country) <u>10801 University Boulevard</u> <u>Manassas, Virginia 20110-2209</u> <u>United States of America</u>	
Date of deposit <u>7 May 1998</u>	Accession Number <u>209852</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	
For receiving Office use only	For International Bureau use only
<input checked="" type="checkbox"/> This sheet was received with the international application  Authorized officer	<input type="checkbox"/> This sheet was received by the International Bureau on:  Authorized officer

*What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
  - (f) a polynucleotide which is a variant of SEQ ID NO:X;
  - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
  - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
  - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
9. A recombinant host cell produced by the method of claim 8.
10. The recombinant host cell of claim 9 comprising vector sequences.
11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;
  - (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;



- (g) a variant of SEQ ID NO:Y;
  - (h) an allelic variant of SEQ ID NO:Y; or
  - (i) a species homologue of the SEQ ID NO:Y.
12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.
13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.
14. A recombinant host cell that expresses the isolated polypeptide of claim 11.
15. A method of making an isolated polypeptide comprising:
- (a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and
  - (b) recovering said polypeptide.
16. The polypeptide produced by claim 15.
17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.
18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:
- (a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and
  - (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.
19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:
- (a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and
  - (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

- (a) contacting the polypeptide of claim 11 with a binding partner; and
- (b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

- (a) expressing SEQ ID NO:X in a cell;
- (b) isolating the supernatant;
- (c) detecting an activity in a biological assay; and
- (d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 20.

<110> Human Genome Sciences, Inc. et al

<120> 110 Human Secreted Proteins

<130> PZ021PCT

<140> PCT/US98/27059

<141> 1998-12-17

<150> 60/068,008

<151> 1998-12-18

<150> 60/068,054

<151> 1998-12-18

<150> 60/068,064

<151> 1998-12-18

<150> 60/068,053

<151> 1998-12-18

<150> 60/068,006

<151> 1998-12-18

<150> 60/068,057

<151> 1998-12-18

<150> 60/068,007

<151> 1998-12-18

<150> 60/070,923

<151> 1998-12-18

<150> 60/068,367

<151> 1998-12-19

<150> 60/068,369

<151> 1998-12-19

<150> 60/068,169

<151> 1998-12-19

<150> 60/068,365

<151> 1998-12-19

<150> 60/068,368

<151> 1998-12-19

<160> 628

<170> PatentIn Ver. 2.0

<210> 1

<211> 733

<212> DNA

SUBSTITUTE SHEET (RULE 26)

<213> Homo sapiens

<400> 1

```

gggatccgga gcccaaattct tctgacaaaa ctacacatg cccaccgtgc ccagcacctg      60
aattcgaggg tgcaccgtca gtcttcctct tcccccaaa acccaaggac accctcatga      120
tctcccgga ccttgaggtc acatgcgtgg tgggtggacgt aagccacgaa gaccctgagg      180
tcaagttcaa ctggtacgtg gacggcgtgg aggtgcataa tgccaagaca aagccgcggg      240
aggagcagta caacagcacg taccgtgtgg tcagcgtcct caccgtcctg caccaggact      300
ggctgaatgg caaggagtac aagtgcgaagg tctccaacaa agccctccca acccccatcg      360
agaaaaccat ctccaaagcc aaagggcagc cccgagaacc acaggtgtac accctgcccc      420
catcccgga tgagctgacc aagaaccagg tcagcctgac ctgcctggtc aaaggcttct      480
atccaagcga catcgccgtg gagtgggaga gcaatgggca gccggagaac aactacaaga      540
ccacgcctcc cgtgctggac tccgacggct ccttcttct ctacagcaag ctaccgtgg      600
acaagagcag gtggcagcag gggaacgtct tctcatgctc cgtgatgcat gaggctctgc      660
acaaccacta cagcgagaag agcctctccc tgtctccggg taaatgagtg cgacggccgc      720
gactctagag gat                                     733

```

<210> 2

<211> 5

<212> PRT

<213> Homo sapiens

<220>

<221> Site

<222> (3)

<223> Xaa equals any of the twenty naturally occurring L-amino acids

<400> 2

Trp Ser Xaa Trp Ser

1 5

<210> 3

<211> 86

<212> DNA

<213> Homo sapiens

<400> 3

```

gcgcctcgag atttccccga aatctagatt tccccgaaat gatttccccg aaatgatttc      60
cccgaatat ctgccatctc aattag                                     86

```

<210> 4

<211> 27

<212> DNA

<213> Homo sapiens

<400> 4

```

gcggcaagct ttttgcaaag cctaggc                                     27

```

<210> 5

<211> 271

<212> DNA

<213> Homo sapiens

<400> 5

3

ctcgagattt	ccccgaaatc	tagatttccc	cgaaatgatt	tccccgaaat	gatttccccg	60
aaatatctgc	catctcaatt	agtcagcaac	catagtcccc	cccctaactc	cgcccatccc	120
gccctaact	cgcccagtt	ccgcccattc	tccgccccat	ggctgactaa	ttttttttat	180
ttatgcagag	gccgaggccg	cctcggcctc	tgagctattc	cagaagtagt	gaggaggcct	240
ttttggaggc	ctaggctttt	gcaaaaagct	t			271

<210> 6  
 <211> 32  
 <212> DNA  
 <213> Homo sapiens

<400> 6  
 gcgctcgagg gatgacagcg atagaacccc gg 32

<210> 7  
 <211> 31  
 <212> DNA  
 <213> Homo sapiens

<400> 7  
 gcgaagcttc ggcactcccc ggatccgcct c 31

<210> 8  
 <211> 12  
 <212> DNA  
 <213> Homo sapiens

<400> 8  
 ggggactttc cc 12

<210> 9  
 <211> 73  
 <212> DNA  
 <213> Homo sapiens

<400> 9  
 gcggcctcga ggggactttc ccggggactt tccggggact ttccgggact ttccatcctg 60  
 ccatctcaat tag 73

<210> 10  
 <211> 256  
 <212> DNA  
 <213> Homo sapiens

<400> 10  
 ctcgagggga ctttcccggg gactttccgg ggactttccg ggactttcca tctgccatct 60  
 caattagtca gcaaccatag tcccggccct aactccgcc atcccggccc taactccgcc 120  
 cagttccgcc cattctccgc cccatggctg actaattttt tttatttatg cagaggccga 180  
 ggccgcctcg gcctctgagc tattccagaa gtatgagga ggcttttttg gaggcctagg 240  
 cttttgcaaa aagctt 256

SUBSTITUTE SHEET (RULE 26)

<210> 11  
 <211> 1271  
 <212> DNA  
 <213> Homo sapiens

<400> 11  
 ggggctgggc cctgctcagg tggctctctc cttgcaggga cgggcatgc tctgcaggct 60  
 gtgctggctg gtctcgtaca gcttggctgt gctgttgctc ggctgcctgc tcttctgag 120  
 gaaggcggcc aagcccgag agacccacg gccaccagc ctttctgggg ctcccccaac 180  
 acccgtcac agccggtgtc caccacaacca cacagtgtct agcgctctc tgtccctgcc 240  
 tagccgtcac cgtctcttct tgacctatcg tcaactgccga aatttctcta tcttgctgga 300  
 gccttcaggc tgttccaagg ataccttctt gctcctggcc atcaagtcac agcctggtea 360  
 cgtggagcga cgtgcggcta tccgcagcac gtggggcagg tggggggatg ggctagggcc 420  
 ggcaactgaag ctggtgttcc tcctaggggt ggcaaggatcc gctccccag cccagctgct 480  
 ggccatgag agtagggagt ttgatgacat cctccagtgg gacttcaactg aggacttctt 540  
 caacctgacg ctcaaggagc tgcaactgca gcgctgggtg gtggctgcct gccccaggc 600  
 ccatttcatg ctaaaggag atgacgatgt ctttgtccac gtccccaacg tgttagagtt 660  
 cctggatggc tgggacccag cccaggacct cctggtggga gatgtcatcc gccaagccct 720  
 gccaacagg aacactaagg tcaaatactt catccacccc tcaatgtaca gggccacca 780  
 ctaccacccc tatgctggtg ggggaggata tgtcatgtcc agagccacag tgcggcgct 840  
 ccaggctatc atggaagatg ctgaactctt cccattgat gatgtctttg tgggtatgtg 900  
 cctgaggagg ctggggctga gccctatgca ccatgctggc ttcaagacat ttggaatccg 960  
 gcggccctg gacccttag acccctgcct gtataggggg ctctgctgg ttcaccgct 1020  
 cagcccccctc gagatgtgga ccatgtgggc actggtgaca gatgaggggc tcaagtgtgc 1080  
 agctggcccc ataccacagc gctgaagggt ggggtgggca acagcctgag agtggactca 1140  
 gtgttgatc tctatcgtga tgcgaaattg atgcctgctg ctctacagaa aatgccaaact 1200  
 tggtttttta actcctctca ccctgttagc tctgattaaa aacactgcaa cccaaaaaaa 1260  
 aaaaaaaaaa a 1271

<210> 12  
 <211> 1451  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (937)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (965)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (975)  
 <223> n equals a,t,g, or c

<400> 12  
 gccagttcac tctcggctg gagacactag gtgtcctgcc cctaactctca ccagagccac 60  
 ctcataaacc ctgcaagctc tgcgaagggc agctggggcac agctgaaagc ccagccacca 120  
 gcccatgtcc tgggtgggac tgggcaggag gggccacctc ctactgctga tcaaccgag 180  
 agccctggct gggatccgtc ttccctcacc aacgggagct cgggccccag ggccctgcc 240  
 acctctgtgc accccacact gcagcaggga gcacctgca ggaggaaactg ggcacctgc 300  
 aggggtctgg tggagacgag gatgctacgg aggcagctgc cccatgggac cagtaagagg 360

**SUBSTITUTE SHEET (RULE 26)**

gatcttgggt	gggcttccct	gcagagagga	agccctcagg	agacaccaca	gtaagccatg	420
ctggagacct	ggaggccagg	cccgtcamct	ggggagctgg	ccactaacag	cgggcagaga	480
gcctcccagg	acagccagca	cagcccccca	cacgtccgag	cccacctcct	catttccccg	540
cttcccgcgt	tcccaagcat	gggaggacct	gccggacgca	gcgcaccaty	tytcctaaca	600
gagaccaagt	ctgagcttca	acgcttgccg	agacgacagg	cacgtgcaag	cytytcytyt	660
ccagcaggag	agcccgagc	aggacacagc	gaytctttca	actgcgtccc	gacaaacggt	720
cagccccttc	gctcctgcag	tttatccaag	ctcaggagga	gctttctaaa	gagaacacag	780
ggagacagct	ggctgccaga	gaagcagtcc	tggctctgga	aggctccacc	cagctaacag	840
ggcctgtgac	acaggtggca	gccagcaaga	cccactgcag	cggcatggcc	ctcacagcct	900
ccccctgtccc	tgtcctggga	gcagccccgg	caaagantcc	taccagaac	aytccaggtc	960
agganggcag	ggccttgarg	aargtgarga	catcctggar	arctgtggcc	accaaggtgc	1020
tgcacggcct	ggaggtctcc	acacatctgg	gcaaacggaa	gctttctggg	aggagctggc	1080
tcccaggccc	tgtctccac	gccaccccat	cacagtcgca	cacacagaca	ggctcccaga	1140
ttgtccaccc	tccacaggga	gaagtcaggg	aggtggggcag	gggacggggg	cagccaccgg	1200
ctcagcctgt	gcacgcccac	ccctcccagc	agcacccttc	tccggcccac	ctggctggcc	1260
tgagtctgtg	gactggcact	gcctgataga	actttcagta	ccttcagtgc	ccaaagggcg	1320
cgacgactag	cccttaaaaag	aggctggagc	ccctgaggaa	gcgggtcttt	aggcaaaacc	1380
accgcacggc	agagcagcca	caggacagat	caatgtcttc	tgcgggaaaa	aaaaaaaaaa	1440
aaaaactcgt	a					1451

&lt;210&gt; 13

&lt;211&gt; 2317

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1419)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2165)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 13

cctgcagcta	ccgtccgcaa	ttcccgtctg	acccacgcgt	ccgggcgact	tgcacgtct	60
tcaacatgaa	gatagccaca	gtgtcagtgc	ttctgccctt	ggctctttgc	ctcatacaag	120
atgctgccag	taagaatgaa	gatcaggaaa	tgtgccatga	atttcaggca	tttatgaaaa	180
akggaaaact	gttctgtccc	caggataaga	aattttttca	aagtcttgat	ggaataatgt	240
tcatcaataa	atgtgccacg	tgcaaaatga	tactggaaaa	agaagcaaaa	tcacagaaga	300
gggcccaggca	tttagcaaga	gctcccaagg	ctactgcccc	aacagagctg	aattgtgatg	360
attttaaaaa	aggagaaaga	gatggggatt	ttatctgtcc	tgattattat	gaagctgttt	420
gtggcacaga	tgggaaaaca	tatgacaaca	gatgtgcact	gtgtgctgag	aatgcgaaaa	480
ccgggtccca	aattggtgta	aaaagtgaag	gggaatgtaa	gagcagtaat	ccagagcagg	540
atgtatgcag	tgtttttcgg	ccctttgtta	gagatggaag	acttggatgc	acaagggaaa	600
atgatcctgt	tcttggctct	gatgggaaga	cgcattggcaa	taagtgtgca	atgtgtgcts	660
agctgttyyy	aaaagaagct	gaaaatgccca	agcgagaggg	tgaactaga	attcgacgaa	720
atgctgaaaa	ggatttttgc	aaggaaatwtg	aaaaacaagt	gagaaatgga	aggctttttt	780
gtacacggga	gagtgatcca	gtccgtggcc	ctgacggcag	gatgcatggc	aacaaatgtg	840
ccctgtgtgc	tgaatttttc	aagcagcggt	tttcagagga	aaacagtaaa	acagatcaaa	900
atttgggaaa	agctgaagaa	aaaactaaag	ttaaaagaga	aattgtgaaa	ctctgcagtc	960
aatatcaaaa	tcaggcaaaag	aatggaatac	ttttctgtac	cagagaaaaat	gaccctattc	1020
gtggctccaga	tgggaaaatg	catggcaact	tgtgttccat	gtgtcaagcc	tacttccaag	1080
cagaaaatga	agaaaaagaa	aaggctgaag	cacgagctag	aaacaaaaga	gaatctggaa	1140
aagcaacctc	atatgcagag	ctttgcagtg	aatatcgaaa	gcttgtgagg	aacggaaaaac	1200

SUBSTITUTE SHEET (RULE 26)

ttgcttgcac	cagagagaac	aatcctatcc	agggcccaga	tgggaaagt	catggcaaca	1260
cctgtcccat	gtgtgaggtc	ttcttccaag	cagaagaaga	agaaaagaaa	aagaaggaag	1320
gtwmtcaag	aaacaaaaga	caatctaaga	gtacagcttc	ctttgwkag	ttgtgtagtg	1380
aatmccgcaa	atccaggaaa	aacggacggc	ttttttgcnc	cagagagaat	gaccccatcc	1440
agggcccaga	tggaaaaatg	catggcaaca	cctgtcccat	gtgtgaggcc	ttctttcaac	1500
aagaagaaag	agcaagagca	aaggctaaaa	gagaagctgc	aaaggaaatc	tgcagtgaat	1560
ttcgggacca	agtgaggaat	ggaacactta	tatgcaccag	ggagcataat	cctgtccctg	1620
gcccagatgg	caaaatgcat	ggaaacaagt	gtgccatgtg	tgccagtgtg	ttcaaacttg	1680
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aagggcgggc	gctctagagg	atccctcgag	1740
gggcccagc	ttacgcgtgc	atgcgacgtc	atagctctct	ccctatagtg	agtcgtatta	1800
taagctaggc	actggccgtc	gttttacaac	gtcgtgactg	ggagatctgc	tagcttggga	1860
tctttgtgaa	ggaaccttac	ttctgtgggtg	tgacataatt	ggacaaacta	cctacagaga	1920
tttaaagctc	taaggtaaat	ataaaatttt	taagtgtata	atgtgttaaa	ctagctgcat	1980
atgcttgcgt	cttgagagtt	ttgcttactg	agtatgattt	atgaaaatat	tatacacagg	2040
agctagtgat	tctaattggt	tgtgtatttt	agattcacag	tcccagggt	catttcaggc	2100
ccctcagtc	tcacagctctg	ttcatgatca	taatcagcca	taccacattt	gtagaggjkt	2160
tactngcytt	aaaaaacctc	ccacacctcc	ccctgaacck	gaacataaaa	atggaakgca	2220
atgggtggtg	taactgggta	atgcagcctt	aataaatggg	ttaccaaatt	aaaggccaat	2280
aggccatcca	ccmaaaattt	tccacccaaa	ttaaaag			2317

&lt;210&gt; 14

&lt;211&gt; 1472

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 14

ggccacgcgt	ccgcgggtacg	gtggtgcggc	tgcggcagca	cagaccagct	gcctacatcc	60
ttgtctccac	cgtgctaacc	ctcatggtgc	cctggcacag	cctggacccc	gactcagcgc	120
ttgcagatgc	cttctaccag	cggggctaca	ggtgggctgg	cttcacgtg	gcagctggct	180
ccatctgcgc	catgaacacc	gtcctgctca	gcctcctctt	ctccctgcca	cgcatgtgtc	240
atgccatggc	cgccgatggg	ctcttcttcc	aggtgtttgc	ccatgtgcac	ccccggacac	300
aggtgcctgt	ggcgggcacc	ctggcggtcg	ggctcctcac	ggccttctctg	gcactgctgc	360
tggaccctga	ctgcgtgggt	cagttcctgt	cccttggcac	actcctggcc	tacacattcg	420
tggccaccag	tatcattgtg	ctgcgtctcc	agaagtcttc	cccgccagc	tccccaggcc	480
cagccagccc	tggccccctg	accaagcagc	agagctcctt	ctcagaccac	ctacagctgg	540
tgggcactgt	acacgcctcc	gtccctgagc	caggggagct	gaagccagcc	ctgaggccct	600
acctgggctt	cttggatggg	tacagccctg	gagcagtggt	gacttgggcg	cttggcggtta	660
tgttggcctc	agccatcacc	ataggctgcg	tgcttgtctt	tgggaactcg	accctgcacc	720
tcccacactg	gggttacatc	ctgctgctcc	tgtctaccag	tgtcatgttt	ctgctcagcc	780
tccttgtcct	gggggctcac	cagcaacagt	atcggaaga	cttatttcag	atccccatgg	840
ttcccctgat	tccagccctg	agcatcgctc	tcaacatctg	cctcatgctg	aaacttagct	900
atctgacctg	ggtgcgcttc	tccatctggc	tgctgatggg	acttgacagt	tatttcggct	960
atggcatccg	gcatagcaag	gagaaccagc	gggagctgcc	agggctgaac	tccacacact	1020
acgtggtatt	ccccaggggc	agcctggagg	agacagtgc	ggctatgcag	ccccccagcc	1080
aggcaccagc	acaggaccct	ggccatatgg	agtagctgat	cagcccacac	ttgcccgcgc	1140
ctcccacacc	tgcttgggag	gccagagagg	ccagacaagc	cgagagcccc	ttctgttgtg	1200
ggcagcctgg	gtttgcaggc	ctgcacaggc	tggggagtcc	tcaggacctt	aggaccttca	1260
tccaggggct	gggcttcggg	tcttcaggag	tgggccttgg	ctggtgctgg	tgccatggac	1320
tctgcccaga	gccttcttgt	ttatgatcag	ctccagctac	ctgggcagtt	gtggtggggg	1380
ggatgggaag	gcccacagcc	caagggatcc	ataataataa	ttgcttggcc	agccatgtgg	1440
cctgctggcg	taaaaaaaaa	aaaaaaaaaa	aa			1472

&lt;210&gt; 15

&lt;211&gt; 1016

&lt;212&gt; DNA

SUBSTITUTE SHEET (RULE 26)



&lt;213&gt; Homo sapiens

&lt;400&gt; 15

ggcacgagct	tccgcggcat	gatttccacc	cagccccggt	ccaccccact	cgttcccttt	60
aagatcctgg	ctctggagtc	ggcagatggg	catggcggct	gcagtggctg	caatgacatt	120
ggccccctacg	gtgagcggga	cgaccagcaa	gtgttcaccc	agaagggtgt	gcccagtgcc	180
agccagctct	tctgtcgtct	ctcatctact	gggcagcggg	tgtgtccgt	gcgtccgtg	240
gacggctcac	ccacgacagc	cttcacagt	ctggagtgcg	agggctcccc	ggcggctcgg	300
ctctcggccc	cggcgctacc	tgctcactgg	ccaggccaac	ggcagcttgg	ccatgtggga	360
cctaaccacc	gccatggacg	gcctcggcca	ggccccctga	ggtggcctga	cggagcaaga	420
gctgatggaa	cagctggaac	actgtgagct	ggccccgcgg	gctcctttca	gctccctcat	480
ggggctgtct	ccccagcccc	tcaccccgca	tctccctcac	cagcctccac	tcagcctcca	540
gcaacacctc	cttgtctggc	caccgtggga	gcccagccc	cccgcaggct	gaggccgggc	600
gccgtgggtg	gggcagcttt	gtggaacgct	gccaggaact	ggtgcggagt	gggccagacc	660
tccgacggcc	acccacacca	gccccgtggc	cctccagcgg	tctcggcact	cccctcacac	720
ctcccaagat	gaagctcaat	gaaacttct	ttttgaacaa	cgcagctgcc	atgatgcctt	780
gggatgcct	ggctcctggg	gactcaggtg	cctccctgat	tcctgtggga	acccggggtt	840
caggccaggg	cctccttgga	ataaatgggt	attgttacta	ggtccccacc	ttccctcttt	900
cttggaagcc	aaagtcaccc	tcccaataa	agtctcact	gcaaaaaaaaa	aaaaaaaaaa	960
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaa	1016

&lt;210&gt; 16

&lt;211&gt; 1239

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 16

cccacgcgtc	cgccccacgc	tccgcccacg	cgtccggctg	cggcgcgatg	gcggcggggc	60
tggcgcggct	cctgttgctc	ctcgggtctc	cggccggcgg	gccccgcggc	gcaggtgcag	120
cgaagatgaa	ggtggtggag	gagcccaacg	cgtttggggg	gaacaacccg	ttcttgcttc	180
aggccagtcg	cctccaggcc	aagagggatc	cttcacccgt	gtctggaccc	gtgcatctct	240
tccgactctc	gggcaagtgc	ttcagcctgg	tggagtccac	gtacaagtat	gagttctgcc	300
cgttccacaa	cgtgaccacg	cacgagcaga	ccttcgcgtg	gaacgcctac	agtgggatcc	360
tcggcatctg	gcacgagtgg	gagatcgcca	acaacacctt	cacgggcacg	tggatgaggg	420
acggtgacgc	ctgccgttcc	cggagccggc	agagcaaggt	ggagctggcg	tgtggaaaaa	480
gcaaccggct	ggcccatgtg	tccgagccga	gcacctgcgt	ctacgcgctg	acgttcgaga	540
ccccctcgt	ctgccacccc	cacgccttgc	tagtgtaccc	aaccctgcca	gaggccctgc	600
agcggcagtg	ggaccaggtg	gagcaggacc	tggccgatga	gctgatcacc	ccccagggcc	660
atgagaagtt	gctgaggaca	ctttttgagg	atgctggcta	cttaaagacc	ccagaagaaa	720
atgaacccac	ccagctggag	ggaggtcctg	acagcttggg	gtttgagacc	ctggaaaact	780
gcaggaaggc	tcataaagaa	ctctcaaagg	agatcaaaaag	gctgaaaagg	ttgctcacc	840
agcacggcat	cccctacacg	aggcccacag	aaacttccaa	cttgagcac	ttgggccacg	900
agacgcccag	agccaagtct	ccagagcagc	tgcggggtga	cccaggactg	cgtgggagtt	960
tgtgaccttg	tgggtgggaga	gcagaggtgg	acgcggccga	gagccctaca	gagaagctgg	1020
ctggtaggac	ccgcagggac	cagctgacca	ggcttgtgct	cagagaagca	gacaaaacaa	1080
agattcsagg	ttttaattaa	ttcccatact	gataaaaata	actccatgaa	ttctgtaaac	1140
cattgcataa	atgctatagt	gtaaaaaaat	ttaaaaaagt	gttaacttta	aacagttcgc	1200
tacaagtaaa	tgattataaa	tactaaaaaa	aaaaaaaaaa			1239

&lt;210&gt; 17

&lt;211&gt; 1405

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

SUBSTITUTE SHEET (RULE 26)

<221> SITE  
 <222> (1403)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1404)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1405)  
 <223> n equals a,t,g, or c

<400> 17  
 gaattcggca cgaggcagct ttggacatgt cgggcctcat agggagcatg ggggtgtggaa 60  
 gtggtggccg tgggctcaaa agctggctgc ttagtttacc agctgtgtga tctcaagcag 120  
 atcaccttct tcttcagagc ctcagtttgc ctgtcgggtca tgccctgcct ggaggccgtg 180  
 gccttgatcc tgcctatcct gctgggttcca gacccacccc ggggagctgc cgagacacag 240  
 ggggaggggg cegtgggagg cttcaggagc agctgggtgtg aggacgtcag atacctgggg 300  
 aaaaactgga gtttcgtgtg gtcgayyctc rgagtgcacg ccatggcctt tgtgactgga 360  
 gccctggggg tctgggcccc caagtttctg ctcgaggcac gcgtgggttca cgggctgcag 420  
 cctccctgct tccaggagcc gtgcagcaac cccgacagcc tgatttttgg ggcactgacc 480  
 atcatgaccg gcgtcattgg ggtcatcttg ggggcagaag ctgcgaggag gtacaagaaa 540  
 gtcatccag gagctgagcc cctcatctgc gcctccagcc tgcttgccac agccccctgc 600  
 ctctacctgg ctctcgtcct ggccccgacc accctgctgg cctcctatgt gttcctgggc 660  
 cttgggggca tgcttctgtc ctgcaactgg gcagtgggtg ccgacatcct gctgtctgtg 720  
 gtggtggcca gatcccgagg gacggcagaa gcacttcaga tcacgggtggk ycacatcctg 780  
 ggaracctgg cagccctatc tcacaggact tatctctagt gtcctgcggg ccargcgccc 840  
 tgactcctat ctgcagcgtc tccgcagcct gsarcararc tycctgtgct gcgcctttgt 900  
 catgcacctg gggggcggtc gcttcctgct gactgcgctg tacctggaga gagacgagac 960  
 ccgggcctgg cagmctgtca cagggacccc agacagcaat gatgtggaca gcaatgacct 1020  
 ggagagacaa ggccctgctt cgggcygtgg cgccctctaca gaggagccct gaggtccctg 1080  
 cctgcactcg tccctgcctgc aagcctcccg ttgggtcccca cagcagcagt gcctcggttc 1140  
 ctctttggct gatcccgagg actccggctg aggcacatct gccacttttg aattcccggtg 1200  
 tggagagctg gcaggacct gtggctgggc tgggaatgga gctgtcagca ctctgcgtgg 1260  
 gaggcctggg cctgtgcctg catcccgctc aaggctgccc cagcctgggg tccccagcct 1320  
 ggctgctgct gggccctgra taaagagagg ccagtacaaa gcccatggat tttgggcctg 1380  
 taaaaaaaaa aaaaaaaaaa aannn 1405

<210> 18  
 <211> 1534  
 <212> DNA  
 <213> Homo sapiens

<400> 18  
 ggcacgagtc aagcaatctg tccacctcag cctcccaaag tgctggaatt acaagcataa 60  
 ccactgcacc tggtaggcac caaattttga atagagaagt taacctgat gaatcaacac 120  
 ttgcttgaat cgtttggttc tccctcctcc ttgttcattg tctttattct gctcatctgg 180  
 atgttgcaaa gatgtaaaaga ttttttcctt tgttgctata gagtagtgct aactccatca 240  
 ttctggcaga agcaccaaca cccagatccc aaaattaagc atcatttgaa gctataactca 300  
 ctgaaataca gttcttctgg gcagaacaac ttcagaaagg acaaacattg gctttctggc 360  
 cacacggaag aggcaattt aataaaggaa gaatggaagt aatgttcaat atcaccagga 420  
 tagcctacct gaaatttctt agtacaaatg agaagcaatg tggcccatgt ggctgtata 480  
 tgagcacgtg gaaagagtga tcttaagttc tccactgatc ctaatcotaat ctgttgatt 540  
 cgcgtgtacc cttaattatc catttaaaaa gctaattatt gctgttagtg gtggtggtgg 600

**SUBSTITUTE SHEET (RULE 26)**

ttgcatgtat	atacgtgttt	caaaaccatc	tgacacagcc	tattcaatgc	tttcctctag	660
ttcatacctc	tgtactgggt	tcctgtgtct	gccacagaaa	ttgccaccaa	ggtagcttaa	720
agtaatggaa	atctgctctc	tcacttttct	ggaggctaca	aatctgcaat	caagggtgtca	780
acaggccatg	ctccctctga	aggctctgcg	gaagaatcct	ttcttgcttc	ttcctagttt	840
tgatggttgc	tgccaatcct	tggcattccc	tggcttggtg	ctgcaacact	ccaatctctg	900
cctcaatcat	cacatgacct	tccttggtga	tctcctttgt	gtctctgtgt	tcaaataatt	960
cttccctttc	tcctgtacat	ataccagtca	ttggatttag	ggttctcttg	atccagtatg	1020
atcttatctc	aactggatta	tatcttcaaa	gaccttattt	gaacgcctta	tttccataac	1080
ggtcacattt	ccaagtacaa	gggttaggac	ttaaaacata	ccttcttggg	ggccacaatt	1140
taacctatta	tatcccacta	cacagtatat	gccatgacag	aacctctttc	atgagaccaa	1200
catgaaatgt	agagtgatat	gattctttat	gtgatcaatg	agtatttgca	gacaatgctg	1260
atgattcccc	tgagaatatc	atggctccct	cccttactta	tttagatata	tggttcaggct	1320
gggtgctgtg	gctcatgcct	ataatcccag	cactttggga	ggggaggaag	gtggttcact	1380
tgagcccagg	agttcaagac	cagcttgggc	aacatggtga	gaccttggtt	ttaccaaaaa	1440
aaagaaagaa	aaatacaaaa	aattagctgg	gtgtggtggc	acatgcctgt	agttcctgct	1500
acttggtggg	ctgggggtggg	agaatcatct	gagc			1534

&lt;210&gt; 19

&lt;211&gt; 1233

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (491)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (493)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (497)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 19

gcagcgtgag	ccaccgtgcc	tggctgggtt	ggggattttc	atattgcatg	tgacagaaaa	60
cctatctgta	actagctaaa	tagccctttc	ccctactgcc	ccccaaaagg	cgggggtttat	120
tgtatcatct	gattcagaaa	tctacgctgg	gcttggttagt	ttgggttcgg	gaacatctgg	180
attccagggtc	tcaaattgaca	tcatogetat	tcatttttct	ctttctctgg	ttctgcccctc	240
caccacgtat	cagctttgtt	ctgtgttggc	ctcagcccca	ttctcaagtt	cacattcagc	300
atgaaaaggc	tgatcacctc	ttccagtcac	tcaagcaaaa	agccccaggt	ttgctgcaat	360
gggctagaat	agtttgactg	taatcacatg	accacctcta	agccaatctc	tgtggggccag	420
gcaccccagt	ggttgggtta	rgcctggatc	ttgtgacaca	ctcctgaaca	catgactcga	480
gggtggggca	nangcanatc	ctcacacaga	actgggtgctg	gattccaggt	ggctgatcaa	540
tcacctgtgt	ccactgtggt	tgtatcaarg	gcgtgtgggt	tgtgctgtgc	tgggcgtaca	600
ctagtgtgtg	cattggccag	ctcggkktgc	cacgttattg	gctttgggat	gctgggtccc	660
aggagtggcg	gtgggaagat	tccacctgct	ttctatggta	ggaaggcagg	tcctgggggtg	720
agggtgagcc	ccgaggggga	ccagtggcca	ctgtggcttg	accgcgaggc	cttgacctga	780
gcattgcagg	catagtttcc	tggccctgta	actgccagat	cgagaccag	tgagacagtg	840
gtttccgtcg	ccagagactg	aaatagtgtg	ttccctgagg	acccctgaag	aagcttggct	900
ttgctacgag	ccagacgtcc	aggtcacaa	tctggctggg	acccaaggct	cagtcccttt	960
ataatagctg	ttgagttcac	ggagtgaagta	caattatgca	tccccactga	ggctcaagag	1020
gtgaagtcac	atagccagtt	agtggtagag	atggggcttg	aacctgggtc	ttcctgactc	1080

SUBSTITUTE SHEET (RULE 26)

10

actcagctac	ctttccatag	aaatagggac	tggaggccgg	gcgagtggtc	tcatgccttt	1140
aatccccagc	acgttggaag	gccgaagagg	gtggatcact	tgagggcagg	agttcaaaca	1200
aaaaaattaa	aaaaaaaaaa	aaaaagggcg	gcc			1233

<210> 20  
 <211> 1090  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (4)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (17)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (47)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1033)  
 <223> n equals a,t,g, or c

<400> 20	
gatntggcga	tacattnaca
atagggaaaag	ctgtacgctg
gcggctcatg	ccccagtat
aaggcgacgc	agctcctcgg
tggtctgtcc	cgctctcact
ggacgtgggc	gctctttacc
ccgcaactac	gccttcgcgt
tgcactgcta	tgcagaagaa
ggcagaggag	gctgaaggag
ggccgtctcc	aaagttccct
aagaggtaac	gcatcagaaa
agaagaccag	tcgttcaaga
ataggaaaata	ttctgatatg
atgataaaaa	gagagttaaa
aaaagaagaa	taagaaaacc
aggacttgtc	agaagctacc
atagggccag	aagcacctat
ggccatgctt	tgnttcccg
agcgattccc	
	60
cagaacagta	gacatgatac
caggaaattccc	gggtcgaccc
ctattccgag	gacatcgtgg
agcagatgtt	cctcttggga
tgagggccmc	aggcctcctt
ctctgggtcg	cgagggcggc
gacctcgat	agtggatata
cggagacta	cgagaaggaa
gtggggaatt	gggagcgctt
ctgacgaaca	tacccagtt
cagattccaa	ctcggaagaa
taagtcgtct	aaaagaaagc
cagagtctga	cacaaattct
aaagaagaa	aaagaaacac
ccagtgtact	aagctgtaaa
cagccaaatg	tggcagatac
tgaaaaacct	ttgaagtatg
agctatggct	gagtatgtta
	1020
	1080
	1090

<210> 21  
 <211> 682  
 <212> DNA  
 <213> Homo sapiens

SUBSTITUTE SHEET (RULE 26)

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (624)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 21

gcaaatatta	attgccattt	actttgaaac	ctaaaatggt	caagattcca	ttttcctcca	60
ggctaataaaa	taataatttg	caatatatag	attgtatttt	gtctttgaaa	cgctgtgagg	120
agatcctctt	aatgtggcat	ggtctgcttc	tatgccttgc	ttctgtgttt	cttgagctcc	180
gtggagatag	gccccctctc	ctggcttctc	tgcttgagcc	acataaaatg	ccacttcaca	240
gctcttccct	ttgaagcctg	atccagtatg	catttgagc	taattactgc	agttgacaca	300
actccatcta	aaagcgtcat	gaaagattct	gtaatcactg	ataagaaaat	gatcttgcaa	360
attattgctg	tgtcctcctt	tattgcctct	ttaccttaac	agtacagttt	acaataatgt	420
aaattttttt	ctaattcttc	aactttaacc	ctagaaattg	tagatgtttt	agcagtgggt	480
atgtgatatt	ggcacaacat	aactatataa	tttgcctaat	attgtggtgc	atacctgtaa	540
tcccagctgc	tcaggagtct	gaggcatgag	aatcacatga	acccaggaga	tggagggtgc	600
ggtgagctga	gagcgagtca	ctgnactcca	gccaggacga	cagagtgaat	ccctgtctca	660
aaaaaaaaaa	aaaaaactcg	ag				682

&lt;210&gt; 22

&lt;211&gt; 770

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 22

ccaatactcc	tttactttct	ttgagttaca	atgttgcatc	tattctgttc	acagccccta	60
ggrctccttt	tcctgctgat	ctttctaggt	cttgactctc	tgccctcgtt	cttgaccgct	120
acccggcttc	agagtccaat	aattatattt	tcaactttgt	cctgtatatg	ctocacttct	180
tggctggaac	tctgttcagt	ttatttctct	actttgaact	atctccacgt	agtgccacct	240
tgtttcctga	tctaaggact	gtgcaactcc	tctccagtcg	gccacatctc	tgaccaaggg	300
attcagtga	aactctgggt	tctcacctga	actttgatct	ttaaccccag	ctgacactag	360
taatggcctt	tgtgatcagg	ctctgaaaga	agactgactt	catgtgaacc	atgtgagcat	420
attgttcata	tatccctcaa	ggtggatcct	tcttttctaa	aaggcatcta	aaaagcaacg	480
gaagttcttt	tgaaaatcag	aggctgcctt	tttggtagca	gttctttcat	ttattctgta	540
gaggatccag	attgagctct	ttataaaata	ttcccttaca	taatgtactg	ggatagtcct	600
aacaatagta	aaccttgatc	cacagatcac	atgtgccatt	ggataaaaat	aaataatgca	660
ragraactca	ataagaccag	cctgttttaa	ggagacatca	taaaaacctc	cattaaaaaa	720
aaaaaaaaaa	aactcgaggg	ggggcccgtg	cccaatcgcc	tgtgatgatc		770

&lt;210&gt; 23

&lt;211&gt; 565

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (10)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (21)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

SUBSTITUTE SHEET (RULE 26)

&lt;221&gt; SITE

&lt;222&gt; (538)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 23

tccccccggn	ctggcaggaa	nttcgscacg	agcagaaagc	aaatcagtgg	ttttctggag	60
tttgacatgg	gggtactaac	aagggaactt	tttgggggtgg	tggaatgct	gtatatatttg	120
attgtgggga	tggttacatg	gttgatgca	tttgcacaaa	cacacttaat	ggtaatgcaa	180
aatgaatata	ttttatttta	tgtaaattat	acotcaaagt	tgaatttttt	taaaaagttt	240
cttttaaaaa	gtaaagacat	ttgtgggtgct	tcttgtaaat	tttactgctg	attatgacct	300
tgttttcttta	gatttgactt	tccaagtttg	aaaagaccag	tttaaaaatg	acttttgctg	360
ggcatgggtgg	catgtgcacg	tagtcttacc	tactcaggag	gctgacgcag	gagggtcctt	420
tgagcccagg	agttggaggc	tacagtgagc	tatgattatg	ccactgcact	ccagcttggg	480
tgaaactgtg	agaccctgtc	tctaaaaaca	aaaaaaaaaa	aaaaaaactc	gaggggggnc	540
cggtagcccaa	ttcgcccaaa	atgga				565

&lt;210&gt; 24

&lt;211&gt; 1356

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 24

ggtcattaag	tcctagtctc	attaattggt	ttcaagcttt	tcgcctacat	tttagactaa	60
ccctgcttat	tcctgtgaat	caagtagtga	tctcctgcag	cttggaagaa	aaaataaggg	120
atgggtaatg	taaaaatctg	gatcaatata	ctgggtcttg	gcaattatcc	tgcaaatctt	180
gccaggtaat	aaaagttagt	aggggtgccc	taatccggaa	gtttctttgt	ttaggaaaat	240
aaaaacaagg	aacttcatag	accccccaa	aggggaattc	tatatcttga	caagtaaaat	300
tttagatgga	aattatctac	aacaccacac	ttacggaaat	tgctatcttc	actctattat	360
ttgcaaaagg	gttatacaca	gtagcacctt	ctaactgaag	tattaaacag	agtttccatt	420
gctgtcgtat	tttgcttaat	tattatcctt	atagcaggga	taatagttac	taacaaaaag	480
gaagcatgaa	agttttacta	tcaactgagcc	tggtaggact	ttttattggg	tttagtgatg	540
cagtttttaa	tgaaacatgc	cgctttttgga	ttaataacct	tagtaaagga	aatttacaga	600
tacttaaaaa	tcaaatccaa	attattgata	ggctcaggaa	aatgccagct	tcagcctgag	660
tggtctacaa	cctctttaat	aaattccagt	cttcttatgg	attggttaac	gccttattaa	720
gccctctctt	gcttatatgt	cttgatttga	tatttgagct	ctgtatatac	aatactataa	780
ctcgaattgt	ttcttctctg	ctagaagcaa	tcaaactcca	aatgggtgctg	taaactgaac	840
cacacatgga	caggccattc	ttccgaggac	ccttagattg	atcccagggg	agccctagct	900
gctattcccc	attcacgccc	ttttcagcag	gaagtagcca	gaaggagtcg	ccgccccaaa	960
tcccctaaca	gcagtttagt	tggtatctcc	acaggaagta	atggtgtagg	agttactaag	1020
aaattatttt	aggcagatag	agaggaaaag	gggtccttgg	gaagttttca	ttttttaaag	1080
catctctgga	aaagtttctt	gtaaagcccc	ggctcttaga	gccaggctgg	caacctttga	1140
tatgcaaatg	taagccatta	gaaaccaggt	ccaccaggc	caggtgtggt	gctcacgcct	1200
gtaatcccaa	cactttggga	agcctaggca	ggtggatcac	ctgaggtcag	gagttcgaga	1260
ccagcctggc	caacatgggt	aaaccccgcc	tctaataaaa	acacaaaaaa	ctaaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaa			1356

&lt;210&gt; 25

&lt;211&gt; 617

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 25

ggcacagcac	agcctgagat	cttggggatc	cctcagccta	acaccacag	acgtcagctg	60
gtggattccc	gctgcatcaa	ggcctaccca	ctgtctccat	gctgggctct	ccctgccttc	120
tgtggctcct	ggccgtgacc	ttcttgggtc	ccagagctca	gcccttggtc	cctcaagact	180

SUBSTITUTE SHEET (RULE 26)

ttgaagaaga	ggagggcagat	gagactgaga	cgggcgtggcc	gcctttgccc	gctgtcccct	240
gcgactacga	ccactgccga	cacctgcagg	tgccctgcaa	ggagctacag	agggctcgcc	300
cgggcgccctg	cctgtgcccc	ggaytctcca	gccccgcccc	gccgcccgcac	ccgccgcgca	360
tgggagaagt	gcgcattgcg	gccgaagagg	gccgcgcagt	ggteccactgg	tgtgccccct	420
tctccccggt	cctccactac	tggctgctgc	tttgggacgg	cagcgagstg	cgcagaaggg	480
gcccgcgcgt	gaacgctacg	gtccgcagag	ccgaactgaa	ggggctgaag	ccagggggca	540
tttatgtcgt	ttgcgtagtg	gccgctaacg	agggcggggc	aagccgcgtg	ccccaggctg	600
gaggagaggg	cctcgag					617

&lt;210&gt; 26

&lt;211&gt; 648

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 26

ggcacgaggt	gttttaaacc	tcagaaacag	atttgagtgt	ttcagtatta	tagaaacagt	60
gatgactatt	catgctctgc	tagtctatgc	ctgcaactcc	aaatgtttgt	ggttcagtat	120
ttcccaccta	cattttctgtt	tggtgacatt	gtcattttta	acaaatatga	ccgagtctag	180
ttttttcttta	aaaggatagt	ttatgagtaa	tcttttaaac	catttccata	ccatctgtat	240
ataaccattt	cggtagagaa	cacactacac	tgaaccctgc	tttagagctg	tgtgttgagc	300
taaaaatata	atttttttaa	aattgactag	caaaatctat	ggccacactg	agaagccttt	360
gaaaatggca	aatacttttc	atcaccaatt	gcccaattca	tctttcttct	gcttctctag	420
ccttgtagca	aaggctacac	agcagcccac	agtccacagt	ctttttggga	aaattggcct	480
gccaccttct	ttaagctcag	tttatttttg	acttactttc	tttgctgtag	ttatgaacct	540
tggggcatta	aaatcccatg	gcaaggagca	taagagatgt	tctcgtagct	ctgcgttgtg	600
tgaatgtcc	atcttagttt	tgttaaaaaa	aaaaaaaaaa	aaaaaaaa		648

&lt;210&gt; 27

&lt;211&gt; 1388

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 27

ggcacgaggt	aagttgcaag	gtacacccac	gggtgattta	tcactcttac	aaagatgata	60
actaatgaag	accgcatcta	gaatgctctt	actggagatg	gtttacagag	cattttttaat	120
catcatactt	agatttatat	taatatctct	tttcaaaact	aattattcca	aactgtgccc	180
tgagatacca	tttggcctca	agttcttttc	tttcgtctgt	attaagggtg	aaataaaaaa	240
gactagtagg	aaaagaaggc	cttattttatg	aagggtgtct	atagctctga	gcttggttagc	300
tacataaaat	gagtaataac	ctaaataagt	aaaactaatg	aagatctaac	tagattactt	360
tgcttaatat	taacatttta	cccgcctccc	gcogtgaac	atttggcaga	tgctctgcag	420
gactcatgag	gacattgggtg	gctacagctg	cttctggcac	tgccccccca	acccccaggt	480
gaggtgaact	tctttacaca	tccagcaagc	tttagttatc	ttcttctccc	atttgagata	540
actgtggcta	caagaatctc	agttaaatca	gatgttttaa	ttaggtgcca	aaaaatctta	600
cagacactga	actaatactt	aaatcaagga	acacttcagt	tctccataaa	atctggtgcc	660
attttccaaa	gaaacagagg	atctttgttt	cacaccctgt	gtactggaat	tgcaacagtg	720
aggcattcta	gctctcacat	gccaatgcga	gtggcattca	ttcttgctca	ctcattttctg	780
cttctcattg	tcacacttgg	aggctctttg	ggggtatgtt	tcagttgatc	tgagaaactg	840
ggtgttacca	atttactaga	gagtttctta	aaatgtatct	gaaacaaaact	attaatgggc	900
attctgtggg	ggtaaaacca	ggcaacgcct	ccctacacta	tctgtccttt	cagagctaag	960
aatctgttat	tttgaattgt	tcacgaagag	tgattctgac	tctgcttcag	tgcaactttt	1020
acaaaccatc	gagcctcatc	aaaggagtga	gttgagctga	ggaattagag	ttaaagaatac	1080
aggtatagtg	ccgggcgtgg	tgctcacgcc	tgtaatccca	acattttggg	aggacaagga	1140
gggtggatca	cctgaggtca	ggagttcgag	accagcctga	ccaacatgga	gaaacctgtg	1200
ctttactaaa	aatacaaaat	tagctggacg	tggtggcaca	tgctgtgat	cacagctact	1260
caggaggctg	aggcaggaga	atcgcttgaa	cccaggaggc	ggaggttgtg	gtgagccgag	1320

SUBSTITUTE SHEET (RULE 26)

atcacgtcac tgcactccag cctgggcaac aagagtgtaaa ttccatctca aaaaaaaaaa 1380  
 aaaaaaaaaa 1388

<210> 28  
 <211> 616  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (17)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (580)  
 <223> n equals a,t,g, or c

<400> 28  
 cmgctrcetra gcaactnagt gggatseccc gggctgcagg aattcggcac gaggagaacg 60  
 gctgcacgtg ggagatgctc cgtggatgtt tgtagaacgc tggcttccgt gtttcctcgt 120  
 tgtggctgtg gtgggtgtggg tctttgcctg tggaccctgt gaagacaaaag aagacagttt 180  
 tggatgggtca agctattttc ttgcttcagg gctccctccc ctgctttttg aagcctcaca 240  
 aaccaggact gtgagggcag gaaggcttgg ggtctttgtg tgctgagcct cattaggggtt 300  
 ttaagaacct ccctccttcc atctctagct tacgagaggg atgattcatt atcttccctc 360  
 ctgaggctgc agtagaagca gacagtctct gcctccctgc ttgcctttcc tccctcccat 420  
 tcaactgttga ttattgccct caagaataac aggttgccca gctactcgag argcttaagt 480  
 gggaggattg cttgacccca ggagttcgag gctgcagtga gctatgatcg cttcactgcg 540  
 ctatagcctg gcagacacag agagacccta tctcaagcan acagacaaaac aaaaaaaaaa 600  
 aaaaaaaaaa ctcgag 616

<210> 29  
 <211> 828  
 <212> DNA  
 <213> Homo sapiens

<400> 29  
 acgagaacac catgctagtg agttcattcc taacagagga gaacttgcatt cttgactaag 60  
 cattagtgat ctcaaaccct ctgcttatga tttttaaact tctgatcttc agaataatttt 120  
 tccatgagct agctctggct ttgtgcatct caaaccttgt ttctctccca tggctgtcat 180  
 acttctgggtg ccctgagatg cagaatttat ttctacttga tacacacatt tgggtattga 240  
 tgtagggtta gtacagcagg taggttgaga atttctggag cctccctccc tccctttgtt 300  
 ctgacctttc cttagtcata tcatcctaga aagatcttcc ctggcttcgt ctaaaacatg 360  
 gcctctcatt tcattctctc cctgacaacc ctgatgtagt tttcatttca ggactcatca 420  
 ccccaacatc ctttcctgtt tacageccat ctcccctgct agaacacaag ctctgagagg 480  
 tgggaaggcct ctattgtggg ttttggcgaa tccccaatct ctagatgggtg tctggcatgt 540  
 gatagagatt caacaaacac ttcaacaaat aatgaataaa gttaaatttt tcagagtga 600  
 atcatgcctc tcccttctc tgcaggggcg gaggtctgtg ctgggttgcg cggettctgc 660  
 agctccagct ccttgtagt agtctggaga atgatggagc tcagtccatt ttaatcccat 720  
 gaacattaaa tgcgtggatg tgtggatgct gggatggatg gatgacgctc ctgacacggc 780  
 agcttgcagg ggattggcga tttccagtaa ggtgtgctaa gactcgag 828

<210> 30  
 <211> 581

**SUBSTITUTE SHEET (RULE 26)**



&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 30

ggcacgagat	tatgggggca	gtttccccc	tgctgttctt	gtgatagtag	ggaattctca	60
tgagagctga	tggttttaag	tgtggcactt	cttcacgctc	tctctcacct	gatgccatgt	120
aagacgtgcc	ttgcttccac	ttcaccttct	gccatgattg	taagtttcct	gaggcctccc	180
cagccagcca	tgtggaactg	tgagtcaatt	aaaccttttt	tgttttataca	ttaccagtc	240
tcaggtagta	tctttatagc	agtgtgagaa	tggaactaatg	cagtcctctt	gttagattgc	300
ctatgtttct	tgactgtcta	ataagatatg	accttcagtg	atacaaatga	tcagagaacc	360
ttgcaaagct	gatgtggtgg	agagagttca	gggagaacca	gagtgggaaca	aaccggagga	420
atgaatatgg	tagagtgcaa	gtgaatttgc	tcctgatttc	cagaagcagt	tgtcagcaac	480
agagtttacc	cttgtatagc	attaaaatag	tctaggacaa	accagaaagt	ccataatgtc	540
ctccagcatt	aagagagcac	ctcgtgccga	attcggcacg	a		581

&lt;210&gt; 31

&lt;211&gt; 789

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 31

ggcacgagcc	tccttccaga	agttccaggg	ttttctttga	tggttgccat	tctgcttaga	60
gaacttccat	tagcctttct	tttgggtggg	tcttctgggtg	acaaattctg	tttcacttcc	120
tctgagaatg	ttttgcttcc	cttttcattc	ctgaaggaca	tttttgctgg	atataagaat	180
tctgggttaa	tggttctttt	cattgtttta	aaaatatatt	gtactttcag	ctgggctcca	240
tggtttctga	tgagaaattc	gctgtcattt	gacttgttta	tgacgagtag	ttaaggcagt	300
ttttgttttag	tggtctttga	aatgttttgt	cttttgtttt	ttggagtttg	attattgtat	360
gtcttagttt	ggatttcttt	gggttcattc	tgtttagggg	ttgcttacct	aagatctgta	420
gatttatgtc	tcttgccaaa	tttggggaact	tttaagccat	cattcgtaga	gtacagtttc	480
aacccacact	tctttctcct	gtcccttcgt	gagttcagtg	actggagttg	ttatagtccc	540
ataggtcccc	aagactgggt	tttttttttt	tttttttttt	ttttgctggg	acatttcttt	600
ctctcttccc	ttcccatccc	gtcccatcca	gtcccgctcc	gtcccatccc	ttcctgtccc	660
ttcccatccc	ttcccgctcc	ttcccatcct	cggagtctct	ccctgtttcc	caggctggag	720
tgcaatggac	gttctcggtc	cactgcaacc	gccgcctccc	tagttogaat	gattcttctg	780
tctcagcct						789

&lt;210&gt; 32

&lt;211&gt; 884

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 32

ggcacgagca	gaatcagggc	actgagctct	actgtaagtg	tgtatttaat	ctccattttt	60
aatttgaccc	atagaaaatt	gtcttttatag	acaaaaaatg	gtcaaagatg	agcaatttca	120
tatctattac	atgcttagtg	ttcactattt	tggggcatct	tgtttctctc	cagggtgcgc	180
attcatctgt	ttttgagttt	aaaaccttgt	atgtgcttaa	gaccaataga	tattctcagt	240
cacttttttag	acattttttgt	cacctcagtt	ttatcagaac	tagaaaaatc	ttccttaaaa	300
acaactgaaa	ccttttcttt	tgagtggttt	cttttgctaa	tgatgaacta	atattaacaa	360
cttaccttct	aataactttg	tctttgtaac	tcagggttaa	aagcattact	ccaccaatct	420
cttttattct	ccattaaaag	atattatttc	ctttataatt	cattttttatt	ccttattttgt	480
ttagtgcctt	aagtctgttt	ctgtacagct	tgagagactgt	aacttggaag	aatgtagata	540
tattcattat	atactatctc	acttaaatgt	gaattctgaa	gaagtttttc	tgagaaaata	600
aattcgtctt	gtttgagtg	tccttctctg	tctgctgtat	tctaccaggg	ggacagagga	660
aagaagggca	gagaaattct	ttcctggatt	tcatgaaatc	actgaggcca	gagcaacaga	720
tcatttaact	tttcttttat	ccgtgtatgt	attactgaaa	aaaatgggtac	ataatagaat	780

SUBSTITUTE SHEET (RULE 26)

ttgagtttat ttttatctga cttcaggtac atagggacac ttctctgttta aaagaactct 840  
gcattgaatt gtgcctttca taaagcaaaa aaaaaaaaaa aaaa 884

&lt;210&gt; 33

&lt;211&gt; 866

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 33

ccacgcgtcc gctccaaaca aacaaaaaat gaactttatt tagatatatt ttacatatga 60  
tgaagtattt ttttgatgta gtagtttttc tcaccttctt tttagtcttc tctttatcca 120  
tttttctttc tgatgaagaa ttccctgtga gtaggaccca gaacataggc ctttgtcatt 180  
tcaacccttc gttctctgaa taggctgttt attggcaaca ttaactggaa acattttatg 240  
tacagcattg gagtctcact ctgtcgcctc agctcactgc aacctccgcc tcttgggttc 300  
aagtgatgtg cactgtatga actgtgagag caagcatatc attataacat tggacaatga 360  
gccaagacag ttctgatgga cttttgaaga gggatttttc aaaagcattt aactcatcat 420  
attaataaaa taaatcctat gatttatggg aaattctgtt ggatcaactt tggaaactgt 480  
ttactataaa ggtagcatgc gtaggcatga atcttgataa gacaagattc tgatccgggg 540  
ttctgagtgg gtccttatat tctgcagagc tgaaccaggt ggaataggag gagagtttgg 600  
gtaaacagtca aacacaacat ccaaaattat gttgaatgca gtggtgagag ctattccctt 660  
taaaactctc tcttggttct tctgactgtg tcaagaatac tgtatttgtt tgggtactgg 720  
ctgggtttttt tttttttttt tttgaaatgc actccagcct gggcgacaag agtgaaactc 780  
tgtctgaaaa gaaagaaaga aagaaaaaga aagaaaggaa agaagggaagg aagaaaaaaa 840  
agaaaagaaa gaaaaaaa aaaaaa 866

&lt;210&gt; 34

&lt;211&gt; 1694

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 34

ccacgcgtcc gggataaaact atttatcata aatttaaagt ttgttattgt gcctttattg 60  
cacatttttt ctcatgcctt ttattataaa atatacttgt tttcaccttg tttttggtca 120  
aatcccagtt actgtccatg taaacatatg gcaacataag aacgtcactt tttttatcct 180  
gcactgtggc atacctgtct ttacaagaga ttctgtctgca cttacatat tcaaatgatgg 240  
cacagttata gagactttat tattcttaat tctttattta gacttaaaca ttatttgttg 300  
ctaattaaaa cagttcatat gatgaacttt ataataaaat atatattttt caggccaggc 360  
atagtggctc atgtctgtaa tcccagcact ccaggagtct gaggcaggcg gaccacttga 420  
gccaagagt ttgagaccag cctggacaac atagggaggt cctgtctcca caaaaaattt 480  
ttaaaaaatt agctaggcat tgtggtgcat tcttgtggtc ccagctactc cggaggctga 540  
ggtaggagga ccacttgagc ccagagggtc aagggcgcag taagtgtgta gactgagact 600  
ctgtctcaat gataataatc ataataataa taaagatatg taccacagag tcaagaatgc 660  
aataagctgt gagagtgaga cctgtctca atgatctta taatagtcac aataaagata 720  
tatattttgc atcagtctgt gggtaattga atgatgatt ctggagttaa actgagcaga 780  
atggtacagt gctgtactgt acagtacata agaatacagga gaggcagctg ttttaagccc 840  
cagtcttatt atgeccattt aatcaaggca tgctggctcc tgtcatgaat tattctgatg 900  
ttgagtaaaa ctgtcatctt caattcatat cactttcaaa tgttacatta taagagaaat 960  
gtaacctggt aagtactaag ttgtagattg ttacagagtc agaaaagttg gatgtaattg 1020  
ctgatattga ttatggtaag tagtctttcc ttgctaaagg aatgccaaat ttttaacaaa 1080  
ttataaagca atcattttta cacaaagatg aaaatacaca tgggtcctca tatgctaaaa 1140  
tgttatatgt ataggtttaa atatctaatt tggaaaagct agacataata ataggttttc 1200  
atccttgtag ctaccaatat caaactatac agttcttgtt agtacctcct ctagggtgtt 1260  
tgagagaggg gcagagaaaa gagctcaaat tcaagttgtt tttatatctg acataatttt 1320  
ttaaaattat aagtcaaatg tttgttaagg gaggggattt aagttattaa ttaaaataag 1380  
attttttagat cttcagattt tacttatcta tctaccttc cattgttatt gagagttggc 1440

SUBSTITUTE SHEET (RULE 26)

tgtattttct	ttttttttta	ctccctggcc	tgattttaaag	atactctgga	acattccaga	1500
ggcactgaat	ttatcattct	aaaagggcca	tgtagtactc	ccaggaggca	tagagctgag	1560
ccactcaatg	tgtcttgagg	ttccaaaatt	ctgattctat	gcttttatga	acatgtaatt	1620
tagcagatgt	tacctaatag	caaaagaaaa	tgctatcttt	accagaatgc	acttaggaaa	1680
aaaaaaaaaa	aaaa					1694

&lt;210&gt; 35

&lt;211&gt; 1215

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 35

ggcacgagct	tttcagtact	ctcacttatt	catgacacag	gaatgaccct	ttactcaaaa	60
ctcttggtgt	tggtcaaagg	tgagcttctt	ttcccttag	tcttagccta	tgtgttgctg	120
ttgtatattg	ttaccaagtt	caactaccta	atgttgaaagc	tctttccaaa	taagatacaa	180
attaaaaggg	gaagcattgc	cagtaacagg	tccctagaga	gcagtgccag	cctgcctgca	240
agaaaaaggg	agaaacttct	taaaaagttt	taagcctggg	caacataagg	agattgtttc	300
tatgaaaaat	aaaaattagc	caggatgggtg	gtgtacacct	gtagtcccag	ctactcggga	360
agatgagatt	ggaggatcac	ttgggcctgg	gaggttgagg	ctacagtga	ctgtgattgt	420
gccactgtac	tccagtctgg	gcgacagtga	aatcttgtct	aaacaaaaac	aatttaactg	480
ggaagcacag	tggtccttga	ggacatttaa	tatcaggaca	aagagcctat	gaatatatca	540
ctgatgtata	taaaccctaa	ggcgttaata	aaagctaact	gtttagtgtt	atccatttaa	600
gggaacagga	ggaattgcat	aactttttaga	ttagtcatag	tggtgcctaa	gggatatgct	660
gtggtatatt	tgtatagcca	gggcacttag	ccttccaacc	aatttatata	ccatgttctt	720
caactgtggg	tgagattttag	cctcaagatt	tgatttacta	tatgtaagta	cattacttga	780
tttctataaa	gaatcctttag	tggaaagatgt	tattctgaat	tatttatcaa	tatgattaat	840
ccagttagaa	attattaatg	atcttccttt	atactataca	taggataact	tttaacttgt	900
cgctacagtt	gttgctctga	ggatcttaat	ttgtttactt	tctaggctac	atgaagctat	960
cttttttaaa	agtgtactct	tcattttcac	tgttattgtg	tacttagcat	aaaaaactca	1020
aatctaggcc	aggtgcagtg	gctcatgcct	gtaatcccag	cactttggga	ggctgaagca	1080
tgcggtacac	ttgagcccag	gagttcaaga	ccagcctggg	caacatggct	aatgaacact	1140
aagcatagtt	ttgatgtctt	tcatatataa	gactttctca	aattcttaaa	aaaaaaaaaa	1200
aaaaaaaaaa	aaaaa					1215

&lt;210&gt; 36

&lt;211&gt; 1794

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1675)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 36

gctcctctag	tgactgctgg	ggtgctagtt	ttcacggtta	ccacagaact	tgagagagga	60
aaatgagaac	ggcaaattaa	aatgccataa	aattccattc	ttacaagcg	ttagtttttt	120
tttctatat	atccactccc	tcacctatct	tatgacatat	gagaaatcct	aacatgggac	180
ttggtgtata	attagcattg	aatgaattta	agttttcttt	ctttctttct	tttcttttat	240
cttctgtggt	ctcctgcaga	atcagtctat	aaaaagggca	tggtaaaaaa	aaatctatct	300
catagcattg	ttgaaaagat	taaatgacat	aataagatga	tgcatatata	gtagctagca	360
ctgtacctga	tgcatattag	gagcttgata	atattactaa	cattatcatc	atcaatgcta	420
tctgagcaaa	gaggctgttc	cttctttcca	gccagagttc	ttctgttgga	taatatctcc	480
agctgtgaac	cccaaccaga	gaccttgaag	cctttatttc	cttttctctc	attggccttt	540
ggctgaagtc	tcctttctgc	agagaacaaa	gtgaccagct	tttgatgaac	tattttcttc	600

SUBSTITUTE SHEET (RULE 26)

ttttatccat	ttccaagtgt	tagccatagt	gaagaagtgt	agctgggtgc	tagcccaact	660
caagaagtga	taatgtwata	tccaacccaa	gataaactca	aggataactt	tcaacacggc	720
tactcaagca	gttcaggggg	gaggcatctt	gtagagagga	gaccaggaag	tcacttggca	780
gctgggccag	cacacagagg	gccattgctc	cagtaaagca	gctaacctcc	atctcttcat	840
caaaccatcc	tagactcagc	actctgcaga	gaaggagcag	atggagggaa	tgtgtggaga	900
gattagataa	gaaggatttc	taatctagt	ggggagcgag	taaaatgtac	agaagtttga	960
gaagccaagc	tcttatgtam	maatccrccc	ccatcactca	acaatcccca	ctgtatgaat	1020
aaagcctgga	aagttttccaa	ttaaaacagt	gcttgtgaat	attggcaagg	ggtatctttg	1080
tgtgcagggg	acattttaaag	ggagaagggt	gaagatacac	ccttgcctty	taggagtaca	1140
ctttctgaga	gcttgttcac	caggctgtga	gtttctcagt	ctatctgttt	ctcagtctgt	1200
tagtaatgaa	tgtatctccc	acttagcaca	gcagctagca	catagtagat	gcctaacaaa	1260
tgtgtgttaa	attgaatggt	ggaagtctgt	gtcctgaaag	cttttcttca	catattacaa	1320
ggctttttat	ttgttcaaca	catttttacc	aagttttttt	ctgtgttcca	ggtcttttga	1380
ttagctgttt	ttaagtcaca	gaaatgggtg	tcgggaacaa	aaccagtcaa	aagtcctcat	1440
tctacatcat	ttacactttc	ccttcctata	tttataagtt	ttaatatcag	cttctacaat	1500
aggttccaga	acaagtgggtg	ctcaaggaat	ggagaaaatg	actattccaa	ccctagctgt	1560
aggtgaacca	aaaaccccag	agaaatcaaa	gtgtagttta	aagcagtgtc	tctcaagttg	1620
taatgtgcat	atagatcacc	tggggttggt	attaaaatgc	aaattctaaa	agagncagga	1680
gaatatcttg	agcctgggag	ccagagggtg	cagtgcgcgc	agatcattcc	actgcattcc	1740
agcctgggtg	acacagcgag	actccatctt	aaaaaaaaaa	aaaaaaaaact	cgta	1794

&lt;210&gt; 37

&lt;211&gt; 1174

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 37

ggcacgagca	aaggcaggaa	cttttgtcat	ttttgttctt	caagatgtcc	ccaatgccta	60
cagcagtaag	tgcactgctg	ggcacacagt	aggatttcaa	taaacattca	atgaatgaat	120
gaatgaatga	atgaatgaat	gcattgccaa	accttgcagt	gctgcctttt	gttccctgcc	180
tagctgctgc	ctccccagcc	ggcctggctg	ctccccagag	cagagatgtg	ccttttctctg	240
tgagccctgc	cacacagttg	aacattgggt	agagcccatg	ggccaggggc	agaggcagga	300
acacactcag	ggcagcgtgc	tgccttcctc	ccatccttct	agaggcaaag	ccaccacagt	360
ccattcctgt	tgccaagagc	cttgggggta	ggggaagtag	acaggaattc	accactcctt	420
aaaaccagag	aagtccagag	tcctctgagg	gacagggctc	ttgatgttca	gagggaatca	480
gcctcggcct	ggggaggctg	ggatttgggt	ggtttattct	tcagcatcca	cttcttctcc	540
agggagtctg	tcagtgcacg	gaaaagtgc	tcagaacccc	agagccaggc	ctagtggaca	600
ctaggttctg	cctgtcccat	ggaggggtgt	gtggctttga	cggattagaa	gtcaggactt	660
ggccatcatg	actacatgct	ttaaaaaata	tgagttttct	ttttttgttt	tgggtttttt	720
aaggcgggtg	ggggctactt	tgtgtttagg	ttttacatcc	tttgcaataa	agttccacc	780
atccagcttg	tgcaagtga	aagaggga	ggacttcagt	ggctttgcct	tgtctatata	840
tgggccagag	agaaaaaagg	aagagggtg	ggcgcggtg	ctcacgcctg	tgggtccagt	900
actttgggag	gccgaggtg	gaggatcacc	tgaggtcagt	agttgagacc	agcctggcca	960
acatggcaaa	atcccgctc	tactaaaaat	acaaaaatta	cctgggcgtg	gtgggtgggca	1020
tctgtaatcc	cagctactcc	agaggctgag	gcggggaga	ggcttgaacc	caggaggcat	1080
aggctgccag	tgagccgaga	tcattcccatg	gcactctagc	ctaagggata	gagtgagact	1140
ctgtctcaaa	aaaaaaaaaa	aaaaaaaaaa	aaaa			1174

&lt;210&gt; 38

&lt;211&gt; 1087

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

## SUBSTITUTE SHEET (RULE 26)

&lt;222&gt; (408)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1005)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 38

gtccattctt	accaccactc	tcccaggtaa	atgctccatc	ttctctgtct	tgggttgcat	60
ttggtccac	ctggtcttct	ttcagttaac	tcccttcagt	ccaccaatg	cagtcttttc	120
tctgcagcca	aatttttttc	tatagttcag	gtctgatgat	accactttcc	attttcaaac	180
tccttaaatt	ggcctcggag	acctgtgatg	gccaatcttc	cccaccctg	tgggtcaggt	240
tcctctggtg	tgggcctatc	tcttagtgca	aggtccctg	tagactgtcc	tggggctgcc	300
tgtctaattg	tcttctcatg	aaccaggtag	gcargcagga	accactcatc	tagtctgggg	360
gtgactgggg	ggtgcgaaag	ttttgttaca	ggatgtattt	aacacttnca	ttgtctggct	420
caaaattctc	tttgaagaaa	ctgcctttgt	ttaattctgt	ttawttattt	agaacttacc	480
tagaattgat	ttgcacttgc	aaaaatgctg	ctgttttaga	gtttgctttc	taggctgggt	540
gcttgctctg	tcaccttagc	taaaaactag	gtactggagt	gccttttcta	ttcycttctc	600
ttctttcttt	cttcttcttt	tttttttktt	tttttagtgg	agtctggctt	tgtcgcccg	660
gcttaagtgc	agtggcacga	tcccggctca	ctgcmtcctc	cgcctcccag	gakaagccct	720
tgaacctggg	agggagttgc	ggtgagcgga	gatcgggcca	ttgcactcca	gcccgggcaa	780
cagatcctga	ctcttatcaa	aaaaaaaaaa	aaaagaaaat	taaccttagt	ttaactttct	840
tgctttacca	agtttttaat	ttttaaaatt	ttttactctt	ttgtaatagc	acttagttta	900
aaacacaaca	cattgtacag	ctgtactaaa	tatatcttca	tttatatcat	tgttccatag	960
cttttttctt	ttgtgtgtgt	gtgtgtgtgt	gtatgtgtgg	gtgancagtg	agatccgtct	1020
caaaaaaaaa	ccaaccaaac	aaaaaaaaaa	aaaaaaactc	gtaggggggg	atccggtacc	1080
caattcg						1087

&lt;210&gt; 39

&lt;211&gt; 438

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 39

ggcacgaggc	aaatgccata	tatgtttaga	ccagcctttc	tcaattgtgg	cacttttgcc	60
atttttggcc	agttaaattc	tgttgtgggg	gctgtcctgt	gcattgcagg	atgttttagca	120
gcattctctg	cctctacctc	ctaaatgcc	gtgcaccctc	cctgcagtta	aatgaccca	180
aatgtctcca	gacataggca	aatgttcccc	ggagaacaaa	gtcatctttg	gttgagggaac	240
aaaatgaacc	tatatttaag	atggtctaaa	tgtttctgag	atgggtctaaa	cattttataag	300
cttaacaatt	aacttgaaac	tcttcaggac	tagagaagct	atttgtaaat	ttaaactatct	360
gacgccttga	aacaccttta	tcataaagat	aactaaacca	gtgttctttg	tgaatggcca	420
aaaaaaaaaa	aaaaaaaaaa					438

&lt;210&gt; 40

&lt;211&gt; 734

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 40

gctcgtgccg	ctgctgggca	ctgggagcag	ggggcggcca	aaggcagtg	gtgggcaggt	60
ccatgcctcc	cctggccccc	cagctctgca	gggcagtggt	cctgggttct	atcttctgtc	120
tgctgcaggt	gaagcctctg	aacgggagcc	caggccccaa	agatgggagc	cagacagaga	180
aaacgcctc	tgcagaccag	aatcaagaac	agttcgaaga	gcactttgtg	gcctcctcag	240
tgggtgagat	gtggcaggtg	gtggacatgg	cccagcagga	agaagaccag	tctccaaga	300

SUBSTITUTE SHEET (RULE 26)

20

cggcagctgt	tcacaagcac	tctttccacc	tcagcttctg	cttttagtctg	gccagtgtca	360
tggtttttctc	aggagggcca	ttgagggcga	cattcccaaa	tatccaactc	tgcttcatgc	420
tcactcactg	accctccctc	cctcctgggc	tccagggtcac	aactcccaaa	ggagatgcag	480
gcatggctct	ctgctctga	tcaccatcac	tgtatctcaa	ggttcagcag	cagagatacc	540
agttgccatc	agtgctaact	gactgcctct	ccagggttcgg	agtttcatct	cccagggcca	600
gagacagcag	acccacatcc	ttctctccca	cacctctcct	ggttttgttc	aggacagcag	660
attagaggca	ggaggcaatg	acaataaaat	aacgataaaa	tcttgagaac	aaaaaaaaaa	720
aaaaaaaaact	cgag					734

&lt;210&gt; 41

&lt;211&gt; 1346

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (707)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (998)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 41

gggcagmgya	aacacctctt	cttataaaca	tggccctgct	ttagcaacag	gtctccatta	80
tgaagcaatt	tggatttggg	catcctatca	agttacttaa	aactaaaactc	tgccgtatag	120
tgttttactt	ggtatttttt	gtgtggccac	agtctagtgt	gatcagagaa	gccacacaga	180
cataaattcc	cagtcctgat	tccattgatt	attagatatg	tacttaccct	gaacatctgt	240
aagatgggga	gtaatgatgt	ttaccacaaa	gagttagtgt	aaggattcga	tgagagaact	300
tgcatttgag	ccatctggag	agaatgttag	tttttagcact	aacgatcgtt	tataatactg	360
ctgaggagga	aggactgaca	tcacctgct	tgagaatcta	gtatatgttg	ggcattccyt	420
atccattcat	gagcctgaga	gtacttgaga	wgttttgatc	ttggacactg	ytgtgcagct	480
ttaaggatga	tgagggggaa	atggaaagag	tcytgaaggg	accagtgtca	gagacagata	540
gaaaagggct	gcctgagccg	ggcgagtgga	ctcacgcctg	taatcccaac	actttgggag	600
gccggggcgg	gtggatcacc	tgagggtcgg	agttcgagac	ccgtctaacc	aacatggaga	660
aacctgctt	ctactawaaa	tcacaaaatta	gccagggtatg	gtggcanatg	cctgttacc	720
cagctgcttg	ggaggctgag	gcaggagaat	tgcttgagag	gcagagggtg	ttgtaagccg	780
agwtgcgcc	attgcacgcc	agcctgggca	acaagatkga	aactcccatc	tcaaaaaaag	840
aaaagaaaag	ggctgcctgg	gtccctggga	agaccagct	cccctctgta	tcttagccaa	900
ctccttaagt	gcacacaacc	tcagctgacc	acatagggtg	cagtgatagc	aaatggtgag	960
gtgggtgggg	gtgggggtcc	tgagctcac	tttctgtnc	gcgtaataag	ggagagatca	1020
gagatgcata	gactttttaa	ctggtagcgt	tcttagagat	ggtccttggc	cttctgttgt	1080
tggtgttgtt	tttttctttt	tcttctcttc	cttctctctc	ttcttctctt	ctccttcttt	1140
cttctttttt	ttttcagagt	cttgctctgt	caccaagact	ggagtgaagt	gatgtgatct	1200
cggcttactg	caacctccac	ctcctgaggt	caggagaatt	gcttgagagg	cagagggtgt	1260
tgtaagccga	gatcgcgcca	ttgcacgcca	gcctgggcaa	caagattgaa	actcccatct	1320
caaaaaaaaa	aaaaaaaaaac	tcgtag				1346

&lt;210&gt; 42

&lt;211&gt; 998

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 42

SUBSTITUTE SHEET (RULE 26)

gggtgtcctgc	ccgttttttgt	ccttaatggc	cgtgttgatt	cgataggaat	cgacagatgg	60
gagccaagag	gggacgtaac	cacttgccag	tgtggcccag	tatttatatg	gtagaagctc	120
gggctggcat	tgagatggag	gtgggaaaga	ggaggggtgc	tcgtgtttaa	agattttctct	180
cttgtgtggt	ccttggttat	ggctcccat	ggtgcttgc	gtcccttgg	tcgcctttcc	240
ttgcattttg	ctttttgcct	ttagtccctc	tgcagtcagg	gaccatgttg	gtgactctcg	300
gtcagatggt	cccatatttg	catgtcttgc	tttggcctct	ctggctctag	gttctgtatt	360
gctagtgtgt	ttctgacatt	ttttctcatt	gtttgtgcta	ccttggccaa	gaacctcttt	420
gccttctccc	tcgtttgtaa	cattgagatc	ctagtagcag	ttacttctta	gggttggttag	480
gatgacatga	gatgggagag	ttggccatct	gccacacaag	agttcccttc	tcactgccat	540
cctctgccc	gggtgttccc	cagaatctgc	agggcccat	tggatcatct	gctgtctgca	600
caccttcctc	tctcacctct	tggcacttcc	ctcaaaaaag	agagaagtgg	agcacagtaa	660
ataaacgcc	gcgttttctc	cagttcccag	cacctctccg	gaactggatc	ccccaaactc	720
ctctgtcact	ttctgtgtcc	agcgggcccc	tggggtcctt	cactgtcttc	actctgtctc	780
gcctgtgcgc	ttggccttgg	tgtgttaggg	actgttataa	gagctgctgt	ccgatcccca	840
tattcaatct	cacagcccca	ctttgtgtac	acacaccaga	gccatcttct	taaaatcaga	900
tcctgtcact	cctctcccta	aaatctacca	gtgtgttaat	aaaatcttga	attttttagca	960
taaaaaaaa	aaaaaaaaa	aaaaaaaaa	aaactcga			998

<210> 43  
 <211> 658  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (6)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (15)  
 <223> n equals a,t,g, or c

<400> 43						
acgacnagac	aggtnaccgg	taccggaatt	acccgggtac	gaacccacgc	gtaccgatat	60
ttcgtcattt	aagaatatct	acagatatgc	ctttgatttt	gcaagggata	aagatcagag	120
aagccttgat	attgatactg	ctaaatctat	gttagctctt	ctgcttggga	ggacatggcc	180
actgttttca	gtattttacc	agtaacctga	gcaatcaaag	tatcgtgtta	tgaacaaaga	240
tcaatggtac	aatgtattag	aattcagcag	aacagtccat	gctgatctta	gtaactatga	300
tgaagatggt	gcttggcctg	ttcttcttga	tgaatttgtt	gagtggcaaa	aagtccgtca	360
gacatcatag	caagaactat	gtgaagaaaa	tgcaaacctt	tcaattccca	cgtgtatata	420
agctaattgt	atgaggggga	aaaaaatcca	acgggtgcat	tttcattcat	atgaaagact	480
tctcatagta	cttttttttc	ctttttttta	aggagggttt	tcttgttaca	tgtgatgggc	540
attgagccac	acctcttctt	agactgaata	ttgaagtgtt	tgttttgagt	tatgtttata	600
acatttattt	cagaacaata	aagattcaga	tttgtgacaa	aaaaaaaaa	aaaaaaaaa	658

<210> 44  
 <211> 566  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (68)  
 <223> n equals a,t,g, or c

SUBSTITUTE SHEET (RULE 26)

&lt;400&gt; 44

ggctatttag	gtgccctata	gggaaagctg	gtacgcctgc	aggtmccggt	ccggaattcc	60
cgggtcgnc	cacgcgtccg	gtcagagaga	aagaactgac	tgaaacgttt	gagatgaaga	120
aagttctcct	cctgatcaca	gccatcttgg	cagtggctgt	tggtttccca	gtctctcaag	180
accaggaacg	agaaaaaaga	agtatcagtg	acagcgatga	attagcttca	gggttttttg	240
tgttccctta	cccatatcca	tttcgcccac	ttccaccaat	tccatttcca	agattttccat	300
ggtttagacg	taatttttct	attccaatac	ctgaatctgc	ccctacaact	ccccttcccta	360
gcgaaaagta	aacaagaagg	aaaagtcacg	ataaacctgg	tcacctgaaa	ttgaaattga	420
gccacttcct	tgaagaatca	aaattcctgt	taataaaaaga	aaaacaaatg	taattgaaat	480
agcacacagc	attctctagt	caatatcttt	agtgatcttc	tttaataaac	atgaaagcaa	540
aaaaaaaaaa	aaaaaagggg	ggccgc				566

&lt;210&gt; 45

&lt;211&gt; 1277

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 45

ggcacgagga	ataatcctag	ctctatctgc	agggttttat	cgtgggtttcc	ctgagtttaa	60
acatgtaaca	ggggagcaga	aggagggata	aatgatttgt	ttatgctcaa	taaagatggt	120
gctgctgttc	tgtcagctaa	cctttgctct	cataacctgc	ataaaactgc	aaagtcttta	180
tttattctcc	taccagcaaa	ttattggaat	tcatagtcac	gtctaatttt	agttagctta	240
gggatggcac	tcaacagggg	tataaaagg	ggagaacatc	tctctcatte	ctgaatttcc	300
aaggctggga	ttacggtaaa	gctgtgggag	atcacagtaa	ttgagggttc	ccaggacact	360
cccatagtgg	gttatgcttc	ccaaatctgg	ctttctaaag	ataatggaat	ctggcctaac	420
ctgtcttcca	atatttgagt	ctcttctata	gcaccctgtg	cagagtgggt	tgtgttcttc	480
acttgataaa	aacttttttg	gggatgcatc	attacagttt	gagccagggt	atttcatctt	540
caggcagctg	tgatagataa	taatacacaa	taaggaaaca	acgaaacact	ttttaatgga	600
ttagtaagtg	ggaatagata	gggcacctgt	catttccattg	ttgtccgcat	gacctgatga	660
ggtagttaac	tgtagtgctc	acttccattt	tcaggccaaa	aactgggggg	ttggaatagt	720
ggagcaactc	cggcagtgcc	acacagctat	gaaatatcag	gtcagaaccc	cagggatact	780
ggatccaaaa	acctgagagg	aagctatata	tcctttaact	tcttcaccca	aagctgaaat	840
ctgtcttcca	gttctatcct	ccaatgacct	aaggaaaaga	atctcttccg	ctacttgctc	900
ttcagataat	taacacaact	ttcaggcctt	cttttgcttt	cttttcaggc	cacaggacac	960
tggttttctg	tggtcgttcc	cccaaccccc	ccaaaccacg	tatttttctc	atttggttta	1020
ttgcagtagc	tctctaaagg	tctgcctgct	ctgttcccta	ttcttcaccc	agcagccaag	1080
gccacgctta	aaatacaagt	caaatacatt	cattcctctg	ctcagaacac	ttctgcagtt	1140
tttctgtgca	cttgaggtaa	aaccttaagt	caacttgaag	attacagaat	tgaattgtct	1200
tactgaaatt	acctttccaa	aatcgatggt	tcccaactcg	tgccgaattc	gatatcaagc	1260
ttatcgatac	cgctcgac					1277

&lt;210&gt; 46

&lt;211&gt; 442

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 46

gaattcggca	cgagggtgta	gtggggggcca	tcctttttcca	tggttgagtg	agggccatcc	60
tttcatgtgg	ctgagaggga	tccatccttt	cctgtggctg	agcgggatcc	attctttccc	120
gtggctgagt	ggaggcccat	ccttaggtac	ctccagttag	cagcctacac	ccttgaggga	180
tggaaaattg	atttgctctc	tcacagactt	ctctggttct	agctttgggc	tatttatgcg	240
tgaagctgct	aagaacattt	cccaaagtgt	atttgtataa	atacctagga	gtaggattcc	300
tagtgtgagg	tcagagtctg	tctagcagca	taaaaacaac	caaaagattt	tctgagcctc	360
agggcacatg	ggtggggccc	agagagcacc	caggaccgaa	gctgcccagg	acctccctcg	420

SUBSTITUTE SHEET (RULE 26)



agggggggccc gtaacccatt cg

442

&lt;210&gt; 47

&lt;211&gt; 890

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (818)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (819)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (829)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (859)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (887)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 47

cttgtaaatg	tttcttttcc	cttaaataca	gataattcat	ttgtattgct	tattttatta	60
tgagctacaa	caaaaggact	tcaggaacaa	gtaatgtatt	agtatgggtc	aagattggtg	120
ataggaactg	tctcaaaagg	atgggtggta	ttttaaatat	aaatagctaa	tgggggtggt	180
aggcctataa	aattaaatgc	cttgataaaa	atccaaaatg	aatgcaaaat	tgttttcact	240
tgtattgact	ttatgttgta	tgattccaat	ctctgttctg	tttggcactt	gtattttaatt	300
cttcaccttt	gtaagacatt	tgtatatgtt	ggatgtgttc	attcaagcta	tttaatatct	360
ggcactgtta	atacacagta	ctttattgta	cagactgttt	tactgtttta	attgtagttc	420
tgtgtacttt	ttttggatgg	ggctggcatg	ttttctttgt	ttcctggcaa	tacgacgtgg	480
gaatttcaat	gcgttttggt	gtagatgcta	acgtgtcaga	atcctttaca	ttcaactttt	540
ctaagaaaag	catttttcagt	cttgtagtgt	gtgcttacag	taactaattt	tgttgaaaat	600
ggtttcaagt	tattcaaatt	tgtacaggac	tgtaaagatt	tgttgacagc	aaaatgttga	660
agaaaaaagc	ttatagaata	aaagctataa	agtatatatt	aggatctgca	aacaatgaag	720
aattatgtaa	tatattgtac	aaatgtaagc	aaaggctctg	aaataaaatg	ccatagtttg	780
tgaaaaaaaa	aaaaaaaaaa	actcgagggg	gggcccgnna	cccaatcgnc	caaaagtggg	840
tcgtattaca	attcactgng	ccgtcgttta	caacgtcgtg	actgggnaaa		890

&lt;210&gt; 48

&lt;211&gt; 737

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

## SUBSTITUTE SHEET (RULE 26)

&lt;221&gt; SITE

&lt;222&gt; (736)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 48

gctgcgagaa	gacgacagaa	gggggagctc	accagcgcca	ccgtccccgg	cgaagttctg	60
cgctggtcgg	cggagtagca	agtggccatg	gggagcctca	gcggtctgcg	cctggcagca	120
ggaagctggt	ttaggttatg	tgaagagat	gtttcctcat	ctctaaggct	taccagaagc	180
tctgatttga	agagaataaa	tggattttgc	acaaaaaccac	aggaaagtcc	cggagctcca	240
tcccgcactt	acaacagagt	gcctttacac	aaacctacgg	attggcagaa	aaagatcctc	300
atatggtcag	gtcgcttcaa	aaaggaagat	gaaatcccag	agactgtctc	gttggagatg	360
cttgatgctg	caaagaacaa	gatgcgagtg	aagatcagct	atctaattgat	tgccctgacg	420
gtggtaggat	gcattctcat	ggttattgag	ggcaagaagg	ctgccccaaag	acacgagact	480
ttaacaagct	tgaacttaga	aaagaaagct	cgtctgaaag	aggaagcagc	tatgaaggcc	540
aaaacagagt	agcagaggta	tccgtgttgg	ctggattttg	aaaatccagg	aattatgtta	600
taacgtgcct	gtattaaaaa	ggatgtggta	tgaggatcca	tttcataaag	tatgatttgc	660
ccaaacctgt	accatttccg	tatttctgct	gtagaagtag	aaataaattt	tcttaataaa	720
aaaaaaaaaa	aaaaaanc					737

&lt;210&gt; 49

&lt;211&gt; 571

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (249)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (548)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 49

tgcacccacg	cgtccggggg	cgtacgcggg	caagatggag	gcgactacgg	ctggtgtggg	60
ccggctagag	gaagaggcgt	tgcggcgaaa	ggaacggctg	aaggccctac	gggagaaaac	120
cgggcgcaag	gtgagaagtg	tggagtgagg	gtcgcagttg	aggcgtccag	cgttcggggg	180
ccgggtcgcg	cttgaggaga	gcaaagggct	aataaggaaa	gacagctgcc	gagggcgcg	240
atgccgggnc	gctaacgcat	gcgcgagaag	acgggcgccc	tcccacgatg	tctggggctg	300
cttggcgtgg	gactcctctg	gcgctggtgc	ggtcgtcgcg	cacgcgcggg	ggtgggcaar	360
gcatggtcag	cgacccgcag	tccatctgac	tctgtcttcc	cgggtgttgc	tcgtgtaggt	420
atctagggct	gcctgtaggt	tcagatgctt	gttggggttag	gcgtgatttg	ttccgttcc	480
ctatggccta	gctgggtctt	aacccccgcc	tctgattctg	agtcagacag	actccccagt	540
tcgggcangc	aattcccttg	gaacaagggc	a			571

&lt;210&gt; 50

&lt;211&gt; 356

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 50

ccacgcgtcc	gtaaaaactcc	acatttgtct	gcatacaggga	aaatgcatgg	gcacacatcc	60
tccttccttc	cctctctact	ctcctccctt	ccttcaggcc	tcttagcatt	gtttgttttc	120
ccatttctga	tactactact	ccatgctgaa	gatttggccat	attactat	tggaacatt	180

## SUBSTITUTE SHEET (RULE 26)

gagtgataga	actcctagaa	aattttgcaa	gaaatgttac	atactgtata	tcaaactctc	240
agattctagt	gttgaaaaag	tagcctatac	tttgctatta	cttatacctg	ctgccataga	300
aaaaaataa	gtttattcat	gacacattta	catttgatca	taaaaaaaaa	aaaaaa	356

<210> 51  
 <211> 913  
 <212> DNA  
 <213> Homo sapiens

<400> 51						
ggcagcaggt	agtgtcctga	aacgcctatg	acagtgcctg	gcccagggtt	gggcacttac	60
tcatgtttatt	cacttattca	tatgtgggaa	tgaacagttc	agccagtata	aaggacaggt	120
gagtacacac	caaggagatc	ttcagggaga	gatcagctgg	ttttcctgtg	aagatgtcag	180
cttcacccct	ccaccgtctc	cctgtgtctc	tggctctgtt	ccctttccag	gctgtgtctg	240
ctgggttcttt	gggtctgcag	ccacctccca	cccccatgaa	gggcaaacc	agcattatgt	300
tacctcctca	gtataaaaag	agagagggtc	tcaaaaaaaa	aaaaaaaaaa	atccaaaaag	360
ttgtctttgt	cagctttggg	agggcagact	ccatagttgg	agatgggctt	ccaaccaacc	420
aaggagataa	atgccagagg	gagcgaacca	tgccaggctc	aaagcacatc	tctcccaaaa	480
ctccccaggt	ggggaagcag	gccagaggct	ccacaaacc	ctcaggagg	cctggagtc	540
agatgctgta	ctccagtatc	taaacaatca	ctgaatctta	aagctgacag	gttcaaagct	600
cttacttttg	gccgagcgca	gtggcttacg	cctgtaatcc	aggcactttc	ggagctgagg	660
tgggtggatc	acctgaggtc	aggagtgtga	gaccaaccta	gccaacatgg	tgaaaaccca	720
tctctactaa	aaatacaaaa	attagctggg	cgtgttgaca	cgtgcctgta	atcccagcta	780
ctcggtaggc	tgaggcagaa	gaatcgcttg	aaccaggag	gcagagggtg	cagtgaagctg	840
agatcatgcc	actgcactcc	agcgtgggtg	acagagtgg	actcccgtct	tgggaaaaaa	900
aaaaaaaaaa	aaa					913

<210> 52  
 <211> 1356  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (1231)  
 <223> n equals a,t,g, or c

<400> 52						
cccacgcgtc	cgaaagaatg	ttgtggctgc	tcttttttct	ggtgactgcc	attcatgctg	60
aactctgtca	accagggtgc	gaaaatgctt	ttaaagttag	acttagtata	agaacagctc	120
tgggagataa	agcatatgcc	tgggatacca	atgaagaata	cctcttcaaa	gcgatggtag	180
ctttctccat	gagaaaagtt	cccaacagag	aagcaacaga	aatttcccat	gtcctacttt	240
gcaatgtaac	ccagagggtat	cattctgggt	tgtgggttaca	gacccttcaa	aaaatcacac	300
ccttctctgt	gttgagggtgc	aatcagccat	aagaatgaac	aagaaccgga	tcaacaatgc	360
cttcttttcta	aatgacccaa	ctctggaatt	tttaaaaatc	ccttccacac	ttgcaccacc	420
catggaccca	tctgtgcca	tctggattat	tatatattgg	gtgatatttt	gcatcatcat	480
agttgcaatt	gcactactga	ttttatcagg	gatctggcaa	cgtagaagaa	agaacaaaaga	540
accatctgaa	gtggatgacg	ctgaagataa	gtgtgaaaac	atgatcacia	ttgaaaatgg	600
catccccctc	gatccccctg	acatgaagg	agggcatatt	aatgatgcct	tcatgacaga	660
ggatgagagg	ctcaccctc	tctgaagggc	tgttggttctg	cttctcctca	aaattaaaca	720
tttgtttctg	tgtgactgct	gagcatcctg	aaataccaag	agcagatcat	atattttgtt	780
tcaccattct	tctttttgta	taaattttga	atgtgcttga	aagtgaagg	caatcaatta	840
taccaccaa	caccactgaa	atcataagct	attcacgact	caaaatatc	taaaatatc	900
ttctgacagt	atagtgtata	aatgtgggtc	tgtggtattt	gtagttattg	atttaagcat	960
ttttagaaat	aagatcaggc	atatgtatat	attttcacac	ttcaaagacc	taaggaaaaa	1020

SUBSTITUTE SHEET (RULE 26)

taaattttcc	agtggaggat	acatataata	tggtgtagaa	atcattgaaa	atggatcctt	1080
tttgacgata	acttatatca	ctctgtatat	gactaagtaa	acaaaagtga	gaagtaatta	1140
ttgtaaattg	atggataaaa	ttggaattac	tcataacag	ggtagggatt	tatcctgtta	1200
tcacaccaac	agttgattat	atattttctg	natatcagcc	cctaatagga	caattctatt	1260
tggtgaccat	ttctacaatt	tgtaaaagtc	caatctgtgc	taacttaata	aagtaataat	1320
catccaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaa			1356

&lt;210&gt; 53

&lt;211&gt; 1547

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 53

ggcacgagcg	gctgcgggcg	cgaggtgagg	ggcgcgaggt	tcccagcagg	atgccccggc	60
tctgcaggaa	gctgaagtga	gaggccccga	gaggggcccag	cccgccccgg	gcaggatgac	120
caaggccccg	ctgttccggc	tgtggctggt	gctggggctg	gtgttcatga	tectgtgat	180
catcgtgtac	tgggacagcg	cagcgcccg	cacttctact	tgacacgctc	cttctctagg	240
ccgcacacgg	ggccgcccgt	gcccacgccc	ggcccgagca	ggacagggag	ctcacggccg	300
actccgatgt	cgacgagttt	ctggacaatt	tctcatgctg	gcgtgaagca	gagtgacctt	360
cccagaaagg	agacggagca	gcccgcctgc	ccgggggagc	atggaggaga	gcgtgagagg	420
ctacgactgg	tccccgcg	acgcccggcg	cagcccagac	cagggccggc	agcaggcgga	480
gcggaggaac	gtgctgcggg	gcttctgcgc	caactccagc	ctggccttcc	ccaccaagga	540
gcgcgcattc	gacgacatcc	ccaactcgga	gctgagccac	ctgatcgtgg	acgaccggca	600
cggggccatc	tactgctacg	tgcccaaggt	ggcctgcacc	aactggaagc	gcgtgatgat	660
cgtgctgagc	ggaagcctgc	tgccaccg	tgccgacctc	cgccaccg	tgccgacctc	720
gcgcgagcac	ctaccgaacg	ccagcgcgca	cctgaccttc	aacaagttct	ggcgccgcta	780
cgggaagctc	tcccgccacc	tcatgaaggt	caagctcaag	aagtacacca	agttcctctt	840
cgtgcgcgac	cccttcgtgc	gcctgatctc	cgcttccgc	agcaagttcg	agctggagaa	900
cgaggagttc	taccgcaagt	tcgccgtgcc	catgctgcgg	gtgtacgcca	accacaccag	960
cctgcccgcc	tcggcgcgcg	aggccttccg	cgctggcctc	aaggtgtcct	tcgccaaactt	1020
catccagtag	ctgctggacc	cgcacacgga	gaagctggcg	cccttcaacg	agcactggcg	1080
gcaggtgtac	cgctctgccc	acccgtgcca	gatcgactac	gattcctggg	gaagctggag	1140
actctggagc	aggacgccc	gcagctgctg	cagctactcc	aggtggaccg	gcagtcgct	1200
tccccccgag	ctaccggaac	aggaccgcca	gcagctggga	ggaggactgg	ttcgccaaga	1260
tccccctggc	ctggaggcag	cagctgtata	aactctacga	ggccgacttt	gttctcttcg	1320
gctaccccaa	gcccgaac	ctctccgag	actgaaagct	ttcgcggtgc	ttttctctgc	1380
gtgcctggaa	cctgacgcac	gcgcactcca	gtttttttat	gacctacgat	tttgcaatct	1440
gggcttcttg	ttcactccac	tgctctatc	cattgagtag	tgatcgata	ttgtttttta	1500
agattaatat	atttcaggta	tttaatacga	aaaaaaaaaa	aaaaaa		1547

&lt;210&gt; 54

&lt;211&gt; 1338

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 54

cccacgcgtc	cggttcccc	atctgtctct	caggagcgag	atctgatcgc	tgaatttgcc	60
caagtcacaa	attggtccag	ctgctgcttg	cgtgtctttg	catggcaccc	ccacaccaac	120
aagtttgtag	tggccctgct	agatgactca	gtccgtgtgt	ataatgccag	cagcaccata	180
gtccccctcc	tgaagcaccg	gctgcagcga	aatgtggcgt	ctctggcctg	gaagccccctt	240
agtgcctctg	tcttggtgtg	ggcctgccag	agctgcattc	ttatctggac	cctggaccct	300
acctccttgt	ctaccgacc	ctcttctggc	tgtgcccaag	tgctgtctca	ccctgggcat	360
acacctgtta	ccagcttggc	ctgggcccc	agtggggggc	ggctgtcttc	agcttcaccg	420
tggatgctgc	tatccgggta	tgggatgtct	caacagagac	ctgtgtcccc	cttccctggg	480
tccgaggagg	tggggtgacc	aactgctctg	gtccccagac	ggcagcaaaa	tcctggctac	540

SUBSTITUTE SHEET (RULE 26)

cactccttca	gctgtctttc	gagtctggga	ggcccagatg	tggacttggtg	agaggtggcc	600
tactctatca	gggcgctgtc	agactggctg	ctggagccca	gatggcagcc	gactgctgtt	660
cactgtattg	ggagagccac	tgattttactc	cctgtctttt	ccagaacggt	gtggtgaggg	720
aaaggggtgc	gttggagggtg	caaagtcagc	aacgattgtg	gcagatctgt	ctgagacaac	780
aatacagaca	ccagatgggtg	aggagaggct	tgggggagag	gctcactcca	tggctctggga	840
ccccagtggg	gaacgtcttg	ctgtgcttat	gaaaggaaaag	ccaagggtac	aggatggtaa	900
accagtcate	ctcctttttc	gcactcgaaa	cagccctgtg	tttgagctcc	ttccctgtgg	960
cattatccag	ggggagccag	gagcccagcc	ccagctcate	actttccatc	ttccttcaac	1020
aaaggggccc	tgctcagtg	gggctggtcc	acaggccgaa	ttgcccacat	cccgctgtac	1080
tttgtcaatg	cccagtttcc	acgtttttagc	ccagtgtctg	ggcgggcccc	ggaacccct	1140
gctgggggtg	gaggctctat	tcatgacctg	ccccctcttta	ctgagacatc	cccaacctct	1200
gccccttggg	accctctccc	agggccacca	cctgttctgc	cccactcccc	acattcccac	1260
ctctaagaat	aaataagttt	tccttttgtt	ttccaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa					1338

&lt;210&gt; 55

&lt;211&gt; 2071

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 55

cgccacgagc	caaaacaaaa	atthttgtgca	gtcctgtcat	ttaatcatct	tatcagtgcc	60
ccaaggtgac	ttttttcaag	agctctgttg	ataccatgtg	tgtgtgttat	tttctggtgt	120
ttttacagat	ttgggcccag	cttagtcac	tattggtatg	gatttatcca	ggagctggac	180
tgcaaccggg	aaaggggcat	cctgctcaaa	gcctgtttcc	ccacgaacat	tgtcacctta	240
tgccacagca	tagcttgacc	ctgaagatcc	tggagagaaa	gctgggagga	aaaggtgaat	300
ccggaagcaa	ttttactttc	ctgcaactgt	agatcctggc	aacatctgcc	ctgaacttca	360
gctgaactct	tgtgtcccgt	agtcacacca	ctacctcttt	agacaaacat	atcaagagtt	420
tctgtttttc	ttcctccctt	gctgatgtga	acagccaaga	actacagaca	caaccctctc	480
attatcagca	tttctgtctc	tgtcaaacaa	ttagtttata	gatagtcata	atctttcttt	540
cttcgggatg	gtttatcctt	gtatgctgaa	caaaagaaaa	aatgtgaaga	ctgaaagggtg	600
tgatttttca	attctcctcc	cagttaccga	atcacccctc	ttattttttt	ttccctaaagc	660
ctccgttcac	tgtctctccc	tctccctttc	tctttccatg	ttgcaactcca	cagagaaccc	720
agcattgcac	taccatcgtg	gaataatcta	gcgcacaccc	aggaaagctg	aagccacaaa	780
gtccaaagcc	acctttgtac	tcacctgcag	agctccagaa	gaccttgatg	gcagcctgcc	840
tatgctgtgt	gtttgtctata	ttcaatcttt	acggcttcc	gacttctgtg	acagtaagcc	900
aagtgcacaa	atacacttga	tgagaatttc	ctcttttaaa	aatgttattt	gaacaccaca	960
tatttttagat	ttatcttatt	tgaaagtatt	agttccattg	tgcttggaag	ccacactcct	1020
ttagattggg	ggccgagagg	cgacaaccca	acattgagga	gagtttattt	ttaaaccatgg	1080
ctagtgtgca	gtatgtacgt	gagctagtat	ttttatgagt	cgagtttttt	aaaggcacat	1140
tctgtatact	gcttagtata	tgcattttat	accatgtaat	tataaaacac	tcgagtaagt	1200
tcagcattag	aaatgttttag	ccttgtagta	actgagtggtg	ccagaaataa	acctggagca	1260
atthtttaaa	aagcaaaaata	aaggagattt	ttctattttg	tcactttaat	ttattcactt	1320
ttgtgtactt	ttatgtactg	caaatacagat	ttcagttctaa	agcgaaacat	caataagtta	1380
ataataaacc	tatctttttg	gaagttgaat	attaatctgt	acccaaaacg	tatttagtaa	1440
aatattttgcc	cccgccaccc	tgccatgctg	acataacaac	ttttataatg	ttgaatagat	1500
gatatgggaa	atactaataa	caacaatgta	atthtttgcag	acagctttta	cttatataca	1560
ttggttgatt	tttttcaaaa	gactaaatat	gtcattttata	ctttgtttat	tttctaccaa	1620
agaaggtgtg	taaaaatatg	cctgtgtgct	ttccttttga	aggacacaaa	cctgggtocca	1680
acatgtgtgg	atthtttaactc	tgagtgggtg	gcattaaatc	aaaagagaga	ggcagaagat	1740
gaaatgctaa	agaagggtca	ggcaaacctc	tgtttcagta	taaaattcat	catgcaggct	1800
tctgagtga	atagaatgat	ttgaaaccac	tactgtattg	cctggatata	cacacacaca	1860
cacacacaca	cacactttat	acaaaaatgt	taaaagcagg	tttctgtggc	tgttctaaac	1920
tgttttttct	ttaggaataa	attacattta	tctgcacaga	tgttgaaaat	cctgtttaa	1980
cctgttcaag	gattttgttta	ttttacatta	aacaaattta	ttatgatgaa	cgtgaacaaa	2040
taaattaaaa	aaaaaaaaaa	aaaaaaaaaa	a			2071

SUBSTITUTE SHEET (RULE 26)

<210> 56  
 <211> 1899  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <221> SITE  
 <222> (1439)  
 <223> n equals a,t,g, or c

<400> 56  
 ccacgcgtcc gccacgcgt ccgcccacgc gtccggttacg atttataaaa gcaaactttt 60  
 aatttttcat aatatatcta ttcctaccta ggttttttta atcaaattaa taaaatggta 120  
 ctcccttttg tgttattggt cagaccaaat tttatcagtg tccttcaccc tttattctac 180  
 tcacattggt tatttctata cttaataagt cctgttcaact ctccctctat aatatattac 240  
 aaacctgatc attgtcacta caccocatc attcctgggc tactacaata gccagggtcta 300  
 cctactttcc tttcggcatt tggacaatct gtgtttctgg aagtagccca aatgatcttt 360  
 tttgttctt atattattca gatcctatta ctcttgatc tcaaccctca caagatattc 420  
 ctattatattg cttaataaaa tccaatatcc ctacctgggc ctgtatactc aataaagtc 480  
 aatatcccta ccttggccta caagatctta gatgactggg atctaggaaa cctattcagc 540  
 taatttcttg caattctcta tgtttacctg ctcaagggtt ttaccactcc tcaaatattc 600  
 aaactgggtt tgcattcttt tagttctatc tattcttttg cagatgcaa tcttagtaat 660  
 cataagcagt ttacctctaa attaatgcta aatattgtaa aagggtgtatt ataagttcag 720  
 tctttcattt attctttcaa ccaccaagtg tgcattgaac acatacattg ggtcacacca 780  
 atcaaataag atagactccc aggcacacc gagcatctac tactcaattc atctcaggat 840  
 ctatacaagg ggtcatttta tcagaaacca aagtcttgat gctgtccgaa aaatcacgtt 900  
 ttcacatga tctcctactt ccctagggtg aagtataact ttttaggata tatcatgggt 960  
 gctgacttaa cttttgtatt ttttaaatat actcatgaca agtatcatat aaaacctaac 1020  
 cagcaacttt gcaccagcaa aagtttttca acatttcaat tcttcaaaaa tcaaatgata 1080  
 taatttctta tgtagtaaaa aattcacaca tctgcaaaagc ttgggtttcac taccacctgt 1140  
 taaaaatctta cctttggaag ctatttatga ttgaaaaaca ctttacctca ctcaaaaaga 1200  
 gctggaagtc tctcttcaat ccaatatgca cacagaagac aaaaagctgt atcatctctt 1260  
 gatgatatat ttgaaatcat atggccacgt ctgtccattg tcttcagagt ttctaagtat 1320  
 ttcagaaaaat tatgacttgc actgtagaac tattttaaag aaattccatg gtgcaaacag 1380  
 aaaaactaaa acttttcatg ttaggataat ttattaaaaa tacaacaaa tcttatgtnt 1440  
 acataagaag atagtaacta gcctttttga gagggaaatt tttctctcat aacttctttt 1500  
 ctagtaattt caataaagaa taactgccat tccaacgttt agcccatctc actctcttgt 1560  
 cttcttatgg ccaagtattc aagcttgaaa tttgcagagg aaattcttgt ccgtttttta 1620  
 tatcatgtgg taagcctaatt aaaacatctt ctgaaataat tagcccttaa aaggatagta 1680  
 tcttctacct gacagaggca aatattattg aaaagtttgt acctataag cacattaatc 1740  
 atggagtccct ggaactggat tctgtctaag actgactttt gcttaattaa gttcacagag 1800  
 attttccaca tatttttcca gaacattgca tgtagagata ttgtcgatca atcacataac 1860  
 tagggtcaga aagatgtaac aaggggagaaa aaaaaaaaa 1899

<210> 57  
 <211> 1543  
 <212> DNA  
 <213> Homo sapiens

<400> 57  
 ggcacgagat ttgattctca tgctcctttc aaaagagcat actagtttgg ggtgggttgg 60  
 tattttctta accttagcaa gccagcttat ttcctatgga agcagaactg gaaacagcag 120  
 atgtccacca tgcttataca ggacactaca cactgtctcg acaagccatg ttctttctct 180  
 cctcttcgtg agcactttct ctggtgatga gttagtattg actacttgaa cctcaaaaact 240

**SUBSTITUTE SHEET (RULE 26)**

gggcctctca	cccaaagcca	aatgaagtag	cgtatgccag	gatgatgttt	cttttggggc	300
ggtggcagtg	agactgctaa	gcaggctgcc	ttaggttttg	ctgtggcaat	gctagcagat	360
tgttccctct	ttcaaagggg	caaaaaatc	attttgggtat	gataactgac	tttctattta	420
cagtttcttg	cccccaaaga	caaaccaagt	ggagacacag	cagctgtatt	tgaagaaggt	480
ggtgatgtgg	acgatttagt	aagtactttt	aatatgcacc	tggtgttctg	tgattgaagt	540
cacctgagct	gtaaatacag	ccacaaaggc	tgattatctt	acacttggtg	cttatttgtg	600
ttttaatttc	caatacacca	gaagcttcct	acaccattat	atattgccat	tataaattca	660
atcagatagg	taatttcata	atagaaattc	ctgtgtttca	tggtgtcggc	tatattgttc	720
attcagatta	atcctctccc	ttgaagggtc	gaaaaagact	agggagctat	tccattagta	780
gcaaaatggt	gtaattcact	gaaattgtcg	ttaacaaaa	ataagtaata	caacatggca	840
ttttgtgtgg	gttgacaaat	gaaacaggcc	ttaaaagggc	tacttcttaa	atgttctcaa	900
ttactttaat	gtaaacaaaa	tagaccgata	ggcatttgag	gatttctgga	ccccattaca	960
ccatgttggt	gatgtctggg	aagctgtgta	gtaaatgtct	tttgtatcta	tccttaatgt	1020
ttggaaactt	cccgcttcta	agcttcata	gacaactgac	caacaaacac	tacgtactat	1080
gatgtcaatc	tttttttagag	acattctcat	tactaaaatg	agtggatact	tgaatgttta	1140
actcctaaaa	taatgagcgg	tgaataaatg	agcaagtaca	tgcatgcctt	ccaatgtaga	1200
gtcattttca	ttaaaccctc	tctcaccaga	gaagcagtg	tatgaaattg	gcctgattcc	1260
tttctaagtg	tggtgttctt	gttcacagtt	ggacatgata	taggtcgtgg	atgtatgggg	1320
aatctaagag	agctgccatc	gctgtgatgc	tgggagttct	aacaaaacaa	gttggatgcg	1380
gccattccaag	gggagccaaa	atctcaagaa	attcccagca	ggttacctgg	aggcgatca	1440
tctaattctc	tgtggaatga	atacacacat	atatattaca	agggataatt	tagaccccat	1500
acaagtttat	aaagagtcac	tgttaaaaaa	aaaaaaaaaa	aaa		1543

&lt;210&gt; 58

&lt;211&gt; 1133

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 58

ggcacgagct	ggagcaaaga	tattgtttga	aggagagttt	atggtttttg	attttaaacg	60
ggcaggggtct	tttttccctct	catttttttg	gacaagagag	gccttcgcct	ttatttttac	120
tctccctctt	ctgctgtccc	tgtgcagagg	aaaaatgaag	aattctccca	gaagtgactt	180
gtcaagactt	aaaaaaaaatg	tttttaatgc	atttcttctt	tgtctagtgc	ctcggtttat	240
ctctaacagg	ggctgtccag	tatatcggtc	ctgttaggag	gggagaaaaa	gttcttccaa	300
aggctggaga	agtgaacaag	gagtc aaatt	tattttccca	attcaacttc	ataattatca	360
tttcttttggc	ttcatgctct	cccgtaaact	atgtgggttg	gatccatccc	atctgggtca	420
cttcagttcta	cttcacgtac	ttgaaaaggc	tttccctttac	acttccagga	ccaaacagca	480
acttccctgcc	acacacttcc	accctatcac	tgggagaaat	ccttttcttg	acatgagcct	540
ttgaacctggg	tggggcagaa	agaaccacaa	actccatctc	ccaatagaac	tttgaaattc	600
actcagcttt	tccttttcatg	ctgtttgttg	cctgcttggt	gcactcctcc	tgcccagaaa	660
ctgcaagatt	tttagcttca	cccctttctg	agagtaattg	tatcttttat	cagaatcagt	720
atcagttccc	ctgtattctg	tgcttcacgt	aatttgcaag	actgacctct	tttaagcatt	780
taattcactc	ccagagtcac	ctggtcaggt	tgcaatatga	ggacttctct	gtctcctctg	840
aagcctggga	cactgagctt	acttaataca	ttagatgttc	aaaagaggag	cgttgtttca	900
tctttcaaaa	tgtaggcca	ttactttgag	tataaaatcg	acttattaat	gattagtaat	960
ttttctaaag	tattgggaaa	actttcttat	tttataagat	cttaacaagc	ttaaaaaaga	1020
attttatgac	cagaatccaa	caagagctct	attttggaat	tgtgcccag	ttggtgatgt	1080
ttactctaaa	attaataata	aaactacttg	taagcaaaaa	aaaaaaaaaa	aaa	1133

&lt;210&gt; 59

&lt;211&gt; 1490

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 59

SUBSTITUTE SHEET (RULE 26)

ggcagaagtt	cctctgcgcg	tccgacggcg	acatgggctg	ccccacggcc	ccggaggccg	60
gcagctggcg	ctggggatcc	ctgctcttcg	ctctcttcct	ggctgcgtcc	ctagacatca	120
cggtgcagc	cctgggtacg	gggtgcctga	tcgtagaatc	ctctgcctcc	ccctcatcct	180
gctcctggtc	tacaagcaaa	ggcaggcagc	ctccaaccgc	cgtgcccagg	agctgggtgcg	240
gatggacagc	aacattcaag	ggattgaaaa	ccccggcttt	gaagcctcac	cacctgcccc	300
ggggataccc	gaggccaaag	tcaggcacc	cctgtcctat	gtggcccagc	ggcagccttc	360
tgagctctgg	cggcatctgc	tttcggagcc	cagcaccccc	ctgtctcctc	caggccccgg	420
agacgtcttc	ttcccatccc	tggaccctgt	ccctgactct	ccaaactttg	aggtcatcta	480
kcccwktgg	gggacagtgg	gctgttgtgg	ctgggtctgg	ggcagggtga	tttgagccag	540
ggctggctct	gtgagtggcc	tccttggcct	cggeccctgt	tcctccctc	ctgctctggg	600
ctcagatact	gtgacatccc	agaagcccag	ccccctcaac	cctctgggatg	ctacatgggg	660
atgctggacg	gctcagcccc	tgttccaagg	atthttgggt	gctgagattc	tcccctagag	720
acctgaaatt	caccagctac	agatgccaaa	tgacttacat	cttaagaagt	ctcagaacgt	780
ccagcccttc	agcagctctc	gttctgagac	atgagccttg	ggatgtggca	gcacagtggt	840
gacaagatgg	acactggggc	accctcccag	gcaccagaca	caggggcacgg	tggagagact	900
tctcccccg	ggcagcgcact	tggctcccc	gttttgcccc	aggtctgctc	tctgtcagac	960
ttctcttttg	taccacagtg	gctctggggc	caggcctgcc	tgcccactgg	ccatcgccac	1020
cttccccagc	tgctctctac	cagcagtttc	tctgaagatc	tgtcaacagg	tttaagtcaat	1080
ctggggcttc	cactgcctgc	attccagtcc	ccagagcttg	gtgggtccga	aacgggaagt	1140
acatatgggg	gcatgggtgg	ctccgtgagc	aaatgggtgtc	ttgggcaatc	tgaggccagg	1200
acagatgttg	ccccaccac	tggagatggt	gctgagggag	gtgggtgggg	ccttctggga	1260
aggtgagtgg	agagggggc	ctgcccccg	ccctccccat	ccctactcc	cactgtctcag	1320
cgcgggccat	tgcaaggggtg	ccacacaatg	ttttgtccac	cctgggacac	ttctgagtat	1380
gaagcgggat	gctattaaaa	actacatggg	gaaacaggtg	caaaccctga	aaaaaaaaaa	1440
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa		1490

&lt;210&gt; 60

&lt;211&gt; 1336

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 60

ggcagcaggt	ccagccaagt	tatctaccct	caaattcttg	actaatcatg	tttgtgtctt	60
gggtatttaa	aattacatac	atataatatc	tttttgccaa	aaacaaaagt	cttgcttctt	120
gtcaaatgat	tgctaaagta	gatctttacat	tttttggtat	tatgtatata	tttatacaca	180
cccccaacac	acttagtgat	ttctgttatt	tcctagggag	cacagcttta	aggctatgag	240
atacaactaa	aaggagccca	tctattttgt	tttcagccca	attattgtac	tcacatttca	300
ggggagaatc	tgaaattcct	gtcatgttta	cagcaacaat	ctatcattcc	tggctagctc	360
tcagcctctc	tctccttoca	taggttagaa	ttatgtcatt	ttgttactta	gtggccacgt	420
ctattttctga	gaaagactgg	ttacatttat	gtggcatctc	aggtatcatt	aaggaaaagc	480
cagagcaggg	gtgagcagag	gtcaaaacca	cagacgcagc	agggccattt	gccgcctttg	540
gccgggatca	caaccactgc	agtcctccag	caggtaggac	ttgccaagcc	taaggctccc	600
catccaatct	agacagaggg	gcgctcagag	cagactttgc	cgtagcccat	gtctgggtgag	660
cacaacaggg	aatgaattgg	gcaactccact	cccccgcttc	tctggcccag	ccctgaacta	720
gatgagctgc	atthcatgga	gcccatttta	aaacctcttt	ccttatgact	ttgttactca	780
agtcocagagt	tctctgtgca	cttctgtctag	ataaggagtg	taagccctgc	ccccagcac	840
tggcagcacg	ctggggccctc	cccacacagg	acaccggtga	gttcgggggg	aagctgactc	900
aaatcaacct	tgaaatctca	tgaaaacaaa	atgacttgct	tttttatttg	atagtgtaat	960
atcattttat	aaatttttta	gggtttttct	cgttgtaata	ttgtacagtt	ttgcatggcc	1020
tgggtgtgat	atthtttggt	tagaatataa	tgctgacaaa	tgtggatgga	ggggaagata	1080
ctgcttttagc	ctatcactcc	ttattttatt	ttgtttgggt	ttatgccctc	agtgctcttag	1140
ggaactttat	aagagatcct	ctgctaccac	acaatgatgt	ggattctttt	gcacagaaat	1200
atthtaagggtg	ggatggtaaa	aaatgtcaca	aaagactcct	caccaatact	ttatgttgat	1260
atcacttaat	atthaccaga	ctttgctgta	ttgcaataaa	acagagaact	gttaaaaaaa	1320
aaaaaaaaaa	aaaaaa					1336

SUBSTITUTE SHEET (RULE 26)